Sharp estimation in sup norm with random design

Stéphane Gaiffas

SHARP ESTIMATION IN SUP NORM WITH RANDOM DESIGN

STÉPHANE GAÏFFAS

Laboratoire de Probabilités et Modèles Aléatoires
Université Paris 7 Denis-Diderot
email: gaiffas@math.jussieu.fr

Abstract. In this paper, we study the estimation of a function based on noisy inhomogeneous data (the amount of data can vary on the estimation domain). We consider the model of regression with random design, where the design density is unknown. We construct an asymptotically sharp estimator which converges, for sup norm error loss, with a spatially dependent normalisation which is sensitive to the variations in the local amount of data. This estimator combines both kernel and local polynomial methods, and it does not depend within its construction on the design density. Then, we prove that the normalisation is optimal in an appropriate sense.

1. Introduction

In most cases, the models considered in curve estimation do not allow situations where the data is inhomogeneous, in so far as the amount of data is implied to be constant over space. This is the case in regression with equispaced design and white noise models, for instance. In many situations, the data can happen to be concentrated at some points and to be little elsewhere. In such cases, an estimator shall behave better at a point where there is much data than where there is little data. In this paper, we propose a theoretical study of this phenomenon.

The available data $[(X_i, Y_i), 1 \leqslant i \leqslant n]$ is modeled by

$$Y_i = f(X_i) + \xi_i, \tag{1.1}$$

where ξ_i are i.i.d. centered Gaussian with variance σ^2 and independent of X_i. The design variables X_i are i.i.d. of unknown density μ on $[0, 1]$, which is bounded away from 0 and continuous. We want to recover f. When μ is not the uniform law, the information is spatially inhomogeneous. We are interested in recovering f globally, with sup norm loss $\|g\|_\infty := \sup_{x\in[0,1]} |g(x)|$. An advantage of this norm is that it is exacting: it forces an estimator to behave well at every point simultaneously. A commonly used benchmark for the complexity of estimation over some fixed class Σ is the minimax risk, which is given by

$$\mathcal{R}_n(\Sigma) := \inf_{\widehat{f}_n} \sup_{f\in\Sigma} \mathbf{E}_f^n \big\{ \|\widehat{f}_n - f\|_\infty \big\}, \tag{1.2}$$

where the infimum is taken over all estimators. We say that ψ_n is the minimax convergence rate over Σ if $\mathcal{R}_n(\Sigma) \asymp \psi_n$, where $a_n \asymp b_n$ means $0 < \liminf_n a_n/b_n \leqslant \limsup_n a_n/b_n < +\infty$. In the regression model (1.1) with Σ a Hölder ball with smoothness $s > 0$ and μ positive and bounded, we have $\psi_n = (\log n/n)^{s/(2s+1)}$, see Stone (1982). Thus, in this case, the minimax rate is not sensitive to the variations in the amount of data. Indeed, such global minimax

Date: April 6, 2006.
2000 *Mathematics Subject Classification.* 62G05, 62G08, 62G15.
Key words and phrases. random design, nonparametric regression, sharp estimation, inhomogeneous data.

benchmarks cannot assess the *design-adaptation* property of an estimator. Instead of (1.2), an improvement is to consider the spatially dependent risk

$$\sup_{f\in\Sigma} \mathbf{E}_f^n \Big\{ \sup_{x\in[0,1]} r_n(x)^{-1} |\widehat{f}_n(x) - f(x)| \Big\}$$

of some estimator $\widehat{f}_n$, where $r_n(\cdot) > 0$ is a family of spatially dependent normalisations. If this quantity is bounded as $n \to +\infty$, we say that $r_n(\cdot)$ is an upper bound over Σ. Necessarily, the "optimal" normalisation satisfies $r_n(x) \asymp (\log n/n)^{s/(2s+1)}$ for any x (note that the optimality requires an appropriate definition here). Therefore, in order to exhibit such an optimal normalisation, we need to consider the sharp asymptotics of the minimax risk.

2. Results

If $s, L > 0$, we define the Hölder ball $\Sigma(s, L)$ as the set of all the functions $f : [0,1] \to \mathbb{R}$ such that

$$|f^{(k)}(x) - f^{(k)}(y)| \leqslant L|x-y|^{s-k}, \quad \forall x, y \in [0,1],$$

where $k = \lfloor s \rfloor$ is the largest integer $k < s$. If $Q > 0$, we define $\Sigma^Q(s,L) := \Sigma(s,L) \cap \{f \text{ s.t. } \|f\|_\infty \leqslant Q\}$, and we denote simply $\Sigma := \Sigma^Q(s,L)$ (the constant Q needs not to be known). All along this study, we suppose:

Assumption D. There is $\nu \in (0,1]$ and $\varrho, q > 0$ such that

$$\mu \in \Sigma(\nu, \varrho) \text{ and } \mu(x) \geqslant q, \text{ for all } x \in [0,1].$$

In the following, we consider a continuous, non-negative and nondecreasing loss function $w(\cdot)$ such that $w(x) \leqslant A(1+|x|^b)$ for some $A, b > 0$ (typically a power function). Let us consider

$$r_{n,\mu}(x) := \Big(\frac{\log n}{n\mu(x)} \Big)^{s/(2s+1)}. \tag{2.1}$$

We prove in theorem 1 below that this normalisation is, up to the constants, an upper bound over Σ, and that it is indeed optimal in theorem 2. We denote by $\mathbb{E}^n_{f,\mu}$ the integration with respect to the joint law $\mathbb{P}^n_{f,\mu}$ of the observations (X_i, Y_i), $1 \leqslant i \leqslant n$. The estimator used in theorem 1 does not depend, within its construction, on μ.

Theorem 1 (Upper bound)**.** *Under assumption D, if $\widehat{f}_n$ is the estimator defined in section 4 below, we have for any $s, L > 0$,*

$$\operatorname{limsup}_n \sup_{f\in\Sigma} \mathbb{E}^n_{f,\mu} \Big\{ w\Big(\sup_{x\in[0,1]} r_{n,\mu}(x)^{-1} |\widehat{f}_n(x) - f(x)| \Big) \Big\} \leqslant w(P), \tag{2.2}$$

where

$$P := \sigma^{2s/(2s+1)} L^{1/(2s+1)} \varphi_s(0) \Big(\frac{2}{2s+1} \Big)^{s/(2s+1)} \tag{2.3}$$

and φ_s is defined as the solution of the optimisation problem

$$\varphi_s := \operatorname*{argmax}_{\substack{\varphi\in\Sigma(s,1;\mathbb{R}),\\ \|\varphi\|_2 \leqslant 1}} \varphi(0), \tag{2.4}$$

where $\Sigma(s,L;\mathbb{R})$ is the extension of $\Sigma(s,L)$ to the whole real line.

In the same fashion as in Donoho (1994), the constant P is defined via the solution of an optimisation problem which is connected to optimal recovery. We discuss this result in section 3, where further details about optimal recovery can be found. The next theorem shows that $r_{n,\mu}(\cdot)$ is indeed optimal in an appropriate sense. In what follows, the notation $|I|$ stands for the length of an interval I.

Theorem 2 (Lower bound)**.** *Under assumption D, if $I_n \subset [0,1]$ is any interval such that for some $\varepsilon \in (0,1)$,*

$$|I_n|\, n^{\varepsilon/(2s+1)} \to +\infty \quad \text{as } n \to +\infty, \tag{2.5}$$

we have

$$\liminf_n \inf_{\widehat{f}_n} \sup_{f\in\Sigma} \mathbb{E}^n_{f,\mu}\big\{w\big(\sup_{x\in I_n} r_{n,\mu}(x)^{-1}|\widehat{f}_n(x) - f(x)|\big)\big\} \geqslant w\big((1-\varepsilon)P\big),$$

where P is given by (2.3) and the infimum is taken among all estimators. A consequence is that if I_n is such that (2.5) holds for any $\varepsilon \in (0,1)$, we have

$$\liminf_n \inf_{\widehat{f}_n} \sup_{f\in\Sigma} \mathbb{E}^n_{f,\mu}\big\{w\big(\sup_{x\in I_n} r_{n,\mu}(x)^{-1}|\widehat{f}_n(x) - f(x)|\big)\big\} \geqslant w(P). \tag{2.6}$$

This result says that the normalisation $r_{n,\mu}(\cdot)$ cannot be strongly improved: no normalisation is uniformly better than $r_{n,\mu}(\cdot)$ within a "large" interval. This result is discussed in the following section.

3. Discussion

Literature. When the design is equidistant, that is $X_i = i/n$, we know from Korostelev (1993) the exact asymptotic value of the minimax risk for sup norm error loss. If $\psi_n := (\log n/n)^{s/(2s+1)}$, we have for any $s \in (0,1]$ and $\Sigma = \Sigma(s,L)$

$$\lim_{n\to+\infty} \inf_{\widehat{f}_n} \sup_{f\in\Sigma} \mathbb{E}_f\big\{w(\psi_n^{-1}\|\widehat{f}_n - f\|_\infty)\big\} = w(C),$$

where

$$C := \sigma^{2s/(2s+1)} L^{1/(2s+1)} \Big(\frac{s+1}{2s^2}\Big)^{s/(2s+1)}. \tag{3.1}$$

This result was the first of its kind for sup norm error loss. In the white noise model

$$dY^n_t = f(t)dt + n^{-1/2}dW_t, \quad t \in [0,1], \tag{3.2}$$

where W is a standard Brownian motion, Donoho (1994) extends the result by Korostelev (1993) to any $s > 1$. In this paper, the author makes a link between statistical sup norm estimation and the theory of optimal recovery (see below). It is shown for any $s > 0$ and $\Sigma = \Sigma(s,L)$ that the minimax risk satisfies

$$\lim_{n\to+\infty} \inf_{\widehat{f}_n} \sup_{f\in\Sigma} \mathbf{E}_f\big\{w(\psi_n^{-1}\|\widehat{f}_n - f\|_\infty)\big\} = w(P_1), \tag{3.3}$$

where P_1 is given by (2.3) with $\sigma = 1$. When $s \in (0,1]$, we have $P = C$, see for instance in Leonov (1997). Since the results by Korostelev and Donoho, many other authors worked on the problem of sharp estimation (or testing) in sup norm. On testing, see Lepski and Tsybakov (2000), see Korostelev and Nussbaum (1999) for density estimation and Bertin (2004a) for white noise in an anisotropic setting. The paper by Bertin (2004b) works in the model of regression with random design (1.1). When μ satisfies assumption D and $\Sigma = \Sigma^Q(s,L)$ for $s \in (0,1]$, it is shown that

$$\lim_{n\to+\infty} \inf_{\widehat{f}_n} \sup_{f\in\Sigma} \mathbb{E}^n_{f,\mu}\big\{w(v_{n,\mu}^{-1}\|\widehat{f}_n - f\|_\infty)\big\} = w(C), \tag{3.4}$$

where C is given by (3.1) and $v_{n,\mu} := [\log n/(n \inf_x \mu(x))]^{s/(2s+1)}$. Note that the rate $v_{n,\mu}$ differs from (and is larger than) ψ_n when μ is not uniform. A disappointing fact is that $v_{n,\mu}$ depends on μ via its infimum only, which corresponds to the point in $[0,1]$ where we have the least information. Therefore, this rate does not take into account all the other regions with more data.

As a consequence, the results presented here are extensions of both the papers by Donoho (1994) and Bertin (2004b): our results are stated in the regression model with random design, where the design density is unknown. In particular, we provide the exact asymptotic value of the minimax risk in regression with random design for any $s > 0$, which was known only for $s \in (0, 1]$ beforehand. Nevertheless, the main novelty is, in our sense, the introduction of a spatially dependent normalisation factor for the assessment of an estimator, with an appropriate optimality criterion. The asymptotically sharp minimax framework is considered here only by necessity.

Optimal recovery. The problem of optimal recovery consists in recovering f from

$$y(t) = f(t) + \varepsilon z(t), \quad t \in \mathbb{R}, \tag{3.5}$$

where $\varepsilon > 0$, z is an unknown deterministic function such that $\|z\|_2 \leqslant 1$ and $f \in C(s, L; \mathbb{R}) := \Sigma(s, L; \mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$. The link between this deterministic problem and estimation with sup norm loss in white noise model was made by Donoho (1994), see also Leonov (1999). The minimax risk for the optimal recovery of f at 0 from (3.5) is defined by

$$E_s(\varepsilon, L) := \inf_T \sup_{\substack{f \in C(s,L;\mathbb{R}) \\ \|f-y\|_2 \leqslant \varepsilon}} |T(y) - f(0)|,$$

where $\inf_T$ is taken among all continuous and linear forms on $\mathbb{L}^2(\mathbb{R})$. We know from Micchelli and Rivlin (1977), Arestov (1990) that

$$E_s(\varepsilon, L) = \inf_{K \in \mathbb{L}^2(\mathbb{R})} \Big(\sup_{f \in C(s,L;\mathbb{R})} \Big| \int K(t)(f(t) - f(0)) \Big| + \varepsilon \|K\|_2 \Big) = \sup_{\substack{f \in \Sigma(s,L;\mathbb{R}) \\ \|f\|_2 \leqslant \varepsilon}} f(0).$$

Note that φ_s satisfies $\varphi_s(0) = E_s(1, 1)$. To our knowledge, the function φ_s is known only for $s \in (0, 1] \cup \{2\}$. The kernel K_s for $s \in (0, 1]$ was found by Korostelev (1993) and by Fuller (1961) for $s = 2$. For any $s > 0$, we know from Leonov (1997) that φ_s is well defined and unique, that it is even and compactly supported and that $\|\varphi_s\|_2 = 1$. A renormalisation argument from Donoho (1994) shows that $E_s(\varepsilon, L) = E_s(1, 1)L^{1/(2s+1)}\varepsilon^{2s/(2s+1)}$, thus it suffices to know $E_s(1, 1)$. If we define

$$B(s, L) := \sup_{f \in C(s,L;\mathbb{R})} \Big| \int K_s(t)(f(t) - f(0)) \Big|, \tag{3.6}$$

we have the decomposition $E_s(1, 1) = B(s, 1) + \|K\|_2$, and in particular, if P is given by (2.3) and

$$c_s := \Big(\frac{\sigma}{L}\Big)^{2/(2s+1)} \Big(\frac{2}{2s+1}\Big)^{1/(2s+1)}, \tag{3.7}$$

we have

$$P = Lc_s^s(B(s, 1) + \|K\|_2). \tag{3.8}$$

About theorem 1. We can understand the result of theorem 1 heuristically. Following Brown and Low (1996) and Brown et al. (2002), we can say that an "idealised" statistical experiment which is equivalent (in the sense that the LeCam deficiency goes to 0) to the model (1.1) is given by the heteroscedastic white noise model

$$dY_t^n = f(t)dt + \frac{\sigma}{\sqrt{n\mu(t)}} dB_t, \quad t \in [0, 1], \tag{3.9}$$

where B is a Brownian motion. In view of the result (3.3) by Donoho (1994), which is stated in the model (3.2), and comparing the noise levels in the models (3.2) and (3.9) (with $\sigma = 1$), we can explain informally that our rate $r_{n,\mu}(\cdot)$ comes from the former rate ψ_n where we "replace" n by $n\mu(x)$.

About theorem 2. From Bertin (2004b), we know when $s \in (0, 1]$ that

$$\operatorname{liminf}_n \inf_{\widehat{f}_n} \sup_{f \in \Sigma} \mathbb{E}^n_{f,\mu} \big\{ w(v_{n,\mu}^{-1} \| \widehat{f}_n - f \|_\infty) \big\} \geqslant w(P),$$

where $v_{n,\mu} = [\log n / (n \inf_x \mu(x))]^{s/(2s+1)}$. An immediate consequence is

$$\operatorname{liminf}_n \inf_{\widehat{f}_n} \sup_{f \in \Sigma} \mathbb{E}^n_{f,\mu} \big\{ w\big(\sup_{x \in [0,1]} r_{n,\mu}(x)^{-1} |\widehat{f}_n(x) - f(x)| \big) \big\} \geqslant w(P), \tag{3.10}$$

where it suffices to use the fact that $r_{n,\mu}(x) \leqslant v_{n,\mu}$ for any $x \in [0, 1]$. This entails that $r_{n,\mu}(\cdot)$ is optimal in the classical minimax sense. However, this lower bound is much weaker than the one considered in theorem 2: it does not exclude the existence of another normalisation $\varrho_n(\cdot)$ such that $\varrho_n(x) < r_{n,\mu}(x)$ for "many" x. Therefore, to prove the optimality of $r_{n,\mu}(\cdot)$, we need to localise the lower bound. Indeed, in theorem 2, if we choose $I_n = [0, 1]$ we find back (3.10) and if $I_n = [\bar{x} - (\log n)^\gamma, \bar{x} + (\log n)^\gamma] \cap [0, 1]$ for any $\gamma > 0$ and $\bar{x} \in [0, 1]$ such that $\mu(\bar{x}) \neq \inf_{x \in [0,1]} \mu(x)$, then obviously $v_{n,\mu}$ does not satisfy (2.6).

About assumption D. In assumption D, μ is supposed to be bounded from below, and from above since it is continuous over $[0, 1]$. When μ is vanishing or exploding at a fixed point, we know from Gaïffas (2005) that a wide range of pointwise minimax rates can be achieved, depending on the behaviour of μ at this point. In this case, we expect the optimal normalisation (whenever it exists) to differ from the classical minimax rate ψ_n not only up to the constants, but in order.

Adaptation to the smoothness. The estimator used in theorem 1 depends on the smoothness s of f (see below). In practice, such a parameter is unknown. Therefore, this estimator cannot be used directly: some smoothness-adaptive technique, like Lepski's method (see Lepski et al. (1997)) can be applied. However, this estimator is considered here for theoretical purposes only, and note that even in the white noise model, the problem of sharp adaptive estimation in sup norm over Hölder classes remains open when $s > 1$.

4. Construction of an estimator

The estimator $\widehat{f}_n$ described below uses both kernel and local polynomial methods. Its construction is divided into two parts: first, at some well-chosen discretization points, we use a Nadaraya-Watson estimator with optimal kernel and a design data driven bandwidth. This part of the estimator is used to attain the minimax constant. Then, between the discretization points, the estimator is defined by a Taylor expansion where the derivatives are estimated by local polynomial estimation. We define the empirical design sample distribution

$$\bar{\mu}_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i},$$

where δ is the Dirac mass, and for $h > 0$, $x \in [0, 1]$, we consider the intervals

$$I(x, h) := \begin{cases} [x, x+h] & \text{when } 0 \leqslant x \leqslant 1/2, \\ [x-h, x] & \text{when } 1/2 < x \leqslant 1. \end{cases} \tag{4.1}$$

The choice of non-symmetrical intervals allows to skip boundaries effects. Then we define, when it makes sense, the "bandwidth" at x by

$$H_n(x) := \operatorname*{argmin}_{h \in [0,1]} \Big\{ h \text{ s.t. } h^{2s} \bar{\mu}_n(I(x, h)) \geqslant \log n / n \Big\}, \tag{4.2}$$

which makes the balance between the bias h^s and the variance $[\log n/(n\bar{\mu}_n(I))]^{1/2}$ of the kernel estimator. When the event in (4.2) is empty (which occurs with a very small probability for large n), we take simply $H_n(x) := \max(1-x, x)$. We consider the sequence of points

$$x_j := j\Delta_n, \quad \Delta_n := (\log n)^{-2s/(2s+1)} n^{-1/(2s+1)}, \tag{4.3}$$

for $j \in \mathcal{J}_n := \{0, \ldots, [\Delta_n^{-1}]\}$ where $[a]$ is the integer part of a with $x_{M_n} = 1$, $M_n = |\mathcal{J}_n|$ (the notation $|A|$ stands also for the size of a finite set A). We define $H_n^M := \max_{j \in \mathcal{J}_n} H_n(x_j)$. From Leonov (1997, 1999) we know that the function φ_s defined by (2.4) is even and compactly supported. We denote by $[-T_s, T_s]$ its support and $\tau_n := \min(2c_s T_s H_n^M, \delta_n)$ where $\delta_n = (\log n)^{-1}$ and c_s is given by (3.7).

As usual with the estimation of a function over an interval, there is a boundary correction. We decompose the unit interval into three parts $[0,1] = J_{n,1} \cup J_{n,2} \cup J_{n,3}$ where $J_{n,1} := [0, \tau_n]$, $J_{n,2} := [\tau_n, 1-\tau_n]$ and $J_{n,3} := [1-\tau_n, 1]$. We also define $\mathcal{J}_{a,n} := \{j | x_j \in J_{a,n}\}$ for $a \in \{1,2,3\}$. If φ_s is defined by (2.4), we consider the kernel

$$K_s := \varphi_s / \textstyle\int \varphi_s. \tag{4.4}$$

The "sharp" part of the estimator is defined as follows: at the points x_j, we define $\widehat{f}_n$ by

$$\widehat{f}_n(x_j) := \begin{cases} \dfrac{\dfrac{1}{nH_n(x_j)} \displaystyle\sum_{i=1}^n Y_i K_s\Big(\frac{X_i - x_j}{c_s H_n(x_j)}\Big)}{\max\Big[\delta_n, \dfrac{1}{nH_n(x_j)} \displaystyle\sum_{i=1}^n K_s\Big(\frac{X_i - x_j}{c_s H_n(x_j)}\Big)\Big]} & \text{if } j \in \mathcal{J}_{2,n}, \\ \bar{f}_n(x_j) & \text{if } j \in \mathcal{J}_{1,n} \cup \mathcal{J}_{3,n}. \end{cases} \tag{4.5}$$

This estimator is (up to the correction near the boundaries) a Nadaraya-Watson estimator with the optimal kernel K_s and a bandwidth fitted to the local amount of data. The boundary estimator $\bar{f}_n$ is defined below.

We recall that $k = \lfloor s \rfloor$ where s is the smoothness of the unknown signal f. For any interval $I \subset [0,1]$ such that $\bar{\mu}_n(I) > 0$, we define the inner product

$$\langle f \, , \, g \rangle_I := \frac{1}{\bar{\mu}_n(I)} \int_I f g \, d\bar{\mu}_n,$$

where $\int_I f \, d\bar{\mu}_n = \sum_{X_i \in I} f(X_i)/n$. If $I = I(x,h)$ (see (4.1)), we define $\phi_{I,m}(y) := (y-x)^m$ and we introduce the matrix $\mathbf{X}_I$ and vector $\mathbf{Y}_I$ with entries

$$(\mathbf{X}_I)_{p,q} := \langle \phi_{I,p} \, , \, \phi_{I,q} \rangle_I \quad \text{and} \quad (\mathbf{Y}_I)_p := \langle Y \, , \, \phi_{I,p} \rangle_I,$$

for $0 \leqslant p, q \leqslant k$. Then, we consider

$$\bar{\mathbf{X}}_I := \mathbf{X}_I + \frac{1}{\sqrt{n\bar{\mu}_n(I)}} \mathbf{I}_{k+1} \mathbf{1}_{\Omega_{n,I}^c},$$

where $\Omega_{n,I} := \big\{\lambda(\mathbf{X}_I) > (n\bar{\mu}_n(I))^{-1/2}\big\}$, where $\lambda(M)$ is the smallest eigenvalue of a matrix M and where $\mathbf{I}_{k+1}$ is the identity matrix on $\mathbb{R}^{k+1}$. Note that the correction term in $\bar{\mathbf{X}}_I$ entails $\lambda(\bar{\mathbf{X}}_I) \geqslant (n\bar{\mu}_n(I))^{-1/2}$. When $\bar{\mu}_n(I) > 0$, the solution $\widehat{\theta}_I$ of the system

$$\bar{\mathbf{X}}_I \theta = \mathbf{Y}_I,$$

is well defined. If $\bar{\mu}_n(I) = 0$, we take $\widehat{\theta}_I = 0$. Then, for any $1 \leqslant m \leqslant k$, a natural estimate of $f^{(m)}(x_j)$ is

$$\widetilde{f}_n^{(m)}(x_j) := m!(\widehat{\theta}_{I(x_j, h_n)})_m,$$

where $h_n := (\sigma/L)^{2/(2s+1)}(\log n/n)^{1/(2s+1)}$. The boundary estimator is given by $\bar{f}_n(x_j) := (\widehat{\theta}_{I(x_j,t_n)})_0$, where $t_n := (\sigma/L)^{2/(2s+1)} n^{-1/(2s+1)}$. If we define

$$\Gamma_{n,I} := \big\{ \min_{1 \leqslant m \leqslant k} \|\phi_{I,m}\|_I \geqslant n^{-1/2} \big\}, \tag{4.6}$$

where $\| \cdot \|_I^2 = \langle \cdot \, , \, \cdot \rangle_I$, then for $x \in [x_j, x_{j+1})$, $j \in \mathcal{J}_n$, we take

$$\widehat{f}_n(x) := \widehat{f}_n(x_j) + \Big(\sum_{m=1}^{k} \frac{\widetilde{f}_n^{(m)}(x_j)}{m!} (x - x_j)^m \Big) \mathbf{1}_{\Gamma_{n,I(x_j,h_n)}}. \tag{4.7}$$

5. Proof of theorem 1

The whole section is dedicated to the proof of theorem 1. We denote by $\mathfrak{X}_n$ the sigma-algebra generated by $X_1, \ldots, X_n$ and by $\mathbb{P}_\mu^n$ the joint law of $X_1, \ldots, X_n$. We recall that the discretization points x_j are given by (4.3). We introduce

$$h_{n,\mu}(x) := [\log n/(n\mu(x))]^{1/(2s+1)},$$

and it is convenient to introduce for $j \in \mathcal{J}_n$: $H_j := H_n(x_j)$, $h_j := h_{n,\mu}(x_j)$, $\mu_j := \mu(x_j)$ and $r_j := r_{n,\mu}(x_j)$.

Step 1: approximation by the discretized risk. We introduce the uniform risk

$$\mathcal{E}_{n,f} := \sup_{x \in [0,1]} r_{n,\mu}(x)^{-1} |\widehat{f}_n(x) - f(x)|,$$

and its discretized version $\mathcal{E}_{n,f}^\Delta := \sup_{j \in \mathcal{J}_n} r_j^{-1} |\widehat{f}_n(x_j) - f(x_j)|$. In view of assumption D, we have $\mu(\cdot)^{s/(2s+1)} \in \Sigma(s\nu/(2s+1), \varrho^{s/(2s+1)})$, thus

$$\sup_{x \in [x_j, x_{j+1}]} |r_{n,\mu}(x)^{-1} - r_j^{-1}| \leqslant r_j^{-1} \big(\varrho/q\big)^{s/(2s+1)} \Delta_n^{s\nu/(2s+1)} = o(1) r_j^{-1}. \tag{5.1}$$

Since $f \in \Sigma^Q(s, L)$, writing the Taylor expansion of f at $x \in [x_j, x_{j+1}]$, we obtain:

$$|\widehat{f}_n(x) - f(x)| \leqslant |\widehat{f}_n(x_j) - f(x_j)| + \sum_{m=1}^{k} (\widetilde{f}_n^{(m)}(x_j) - f^{(m)}(x_j)) \frac{(x - x_j)^m}{m!} + L\Delta_n^s,$$

and in view of (5.1):

$$\mathcal{E}_{n,f} \leqslant (1 + o(1)) \Big(\mathcal{E}_{n,f}^\Delta + \max_{j \in \mathcal{J}_n} r_j^{-1} \sum_{m=1}^{k} |\widetilde{f}_n^{(m)}(x_j) - f^{(m)}(x_j)| \frac{\Delta_n^m}{m!} \Big) + O(\delta_n^s),$$

where we recall that $\delta_n = (\log n)^{-1}$. In this step, we need the following lemma, which provides a control over the local polynomial estimator uniform risk. Its proof is given below in the section.

Lemma 1. *There is an event $\mathcal{C}_n \in \mathfrak{X}_n$ such that, under assumption D,*

$$\mathbb{P}_\mu^n \{ \mathcal{C}_n^c \} \leqslant \exp\big(- D_{\mathcal{C}} n^{s/(2s+1)} \big), \tag{5.2}$$

where $D_{\mathcal{C}} > 0$, and a centered Gaussian vector $W \in \mathbb{R}^{(k+1)M_n}$ with

$$\mathbb{E}_{f,\mu}^n \{ W_p^2 \} = 1, \quad 0 \leqslant p \leqslant (k+1)M_n,$$

such that on $\mathcal{C}_n$, one has for any $0 \leqslant m \leqslant k$ and $f \in \Sigma(s, L)$:

$$\max_{j \in \mathcal{J}_n} |\widetilde{f}_n^{(m)}(x_j) - f^{(m)}(x_j)| = O(h_n^{s-m})(1 + (\log n)^{-1/2} W^M), \tag{5.3}$$

where $W^M := \max_{0 \leqslant p \leqslant (k+1)M_n} |W_p|$. *For the estimator near the boundaries, we have on* $\mathcal{C}_n$, *for* $a = 1$ (*the case* $a = 3$ *is similar*):

$$\max_{j \in \mathcal{J}_{1,n}} |\bar{f}_n(x_j) - f(x_j)| = O(t_n^s)(1 + W^{(1)}), \tag{5.4}$$

where $W^{(1)} = \max_{0 \leqslant p \leqslant (k+1)|\mathcal{J}_{1,n}|} |W_p|$.

In view of (5.3), we have on $\mathcal{C}_n$, for any $1 \leqslant m \leqslant k$:

$$\max_{j \in \mathcal{J}_n} r_j^{-1} |\widetilde{f}_n^{(m)}(x_j) - f^{(m)}(x_j)| \Delta_n^m / m! = O(\delta_n^m)(1 + (\log n)^{-1/2} W^M),$$

then,

$$\mathcal{E}_{n,f} \mathbf{1}_{\mathcal{C}_n} \leqslant (1 + o(1)) \mathcal{E}_{n,f}^{\Delta} \mathbf{1}_{\mathcal{C}_n} + O(\delta_n)(1 + \delta_n^{1/2} W^M) + o(1).$$

Let us define the event $\mathcal{W}_n := \{|W^M - \mathbb{E}_{f,\mu}^n\{W^M\}| \leqslant \delta_n^{-1}\}$. Since W is a centered Gaussian vector such that $\mathbb{E}_{f,\mu}^n\{W_p^2\} = 1$ for $0 \leqslant p \leqslant (k+1)M_n$, it is well known (see for instance in Ledoux and Talagrand (1991)) that $\mathbb{E}_{f,\mu}^n\{W^M\} \leqslant [2 \log((k+1)M_n)]^{1/2} = O(\delta_n^{-1/2})$ and

$$\mathbb{P}_{f,\mu}^n\{\mathcal{W}_n^c\} \leqslant 2 \exp(-\delta_n^{-2}/2). \tag{5.5}$$

Thus,

$$\mathcal{E}_{n,f} \mathbf{1}_{\mathcal{C}_n \cap \mathcal{W}_n} \leqslant (1 + o(1)) \mathcal{E}_{n,f}^{\Delta} \mathbf{1}_{\mathcal{C}_n} + o(1). \tag{5.6}$$

Step 2: some events study. In what follows, it is convenient to write K instead of K_s, and to introduce

$$K_{ij} := K[(X_i - x_j)/(c_s h_j)], \quad \bar{K}_{ij} := K[(X_i - x_j)/(c_s H_j)],$$

and $q_j := n c_s h_j \mu_j$, $\bar{q}_j := n c_s H_j \mu_j$, where c_s is given by (3.7). We introduce also

$$\bar{Q}_j := \textstyle\sum_{i=1}^n \bar{K}_{ij}, \quad Q_j := \sum_{i=1}^n K_{ij}, \quad S_j := \sum_{i=1}^n \bar{K}_{ij}^2,$$

and the events

$$\begin{aligned} \mathrm{A}_{n,j} &:= \big\{ \big| \bar{Q}_j / \bar{q}_j - 1 \big| \leqslant L_A \delta_n^{\min(s,1)} \big\}, \quad \mathrm{B}_{n,j} := \big\{ \big| Q_j / q_j - 1 \big| \leqslant \delta_n \big\}, \\ \mathrm{C}_{n,j} &:= \big\{ |H_j / h_j - 1| \leqslant \delta_n \big\}, \quad \mathrm{E}_{n,j} := \big\{ \big| S_j / q_j - \|K\|_2^2 \big| \leqslant L_E \delta_n^{\min(s,1)} \big\}, \\ \mathcal{B}_n &:= \textstyle\bigcap_{j \in \mathcal{J}_{2,n}} \big(\mathrm{A}_{n,j} \cap \mathrm{B}_{n,j} \cap \mathrm{E}_{n,j} \big) \cap \bigcap_{j \in \mathcal{J}_n} \mathrm{C}_{n,j}, \end{aligned} \tag{5.7}$$

where L_A and L_E are some fixed positive constants, $\delta_n = (\log n)^{-1}$, and the sets of indices $\mathcal{J}_{a,n}$ are defined in section 4. In this step, we control the probabilities of these events.

For $j \in \mathcal{J}_{2,n}$, we consider the sequence of i.i.d variables $\zeta_{ij} := K_{ij} - \mathbb{E}_\mu^n\{K_{ij}\}$, $1 \leqslant i \leqslant n$. Since $\mu \in \Sigma_q(\nu, \varrho)$ and $\int K = 1$, we have for n large enough $|\mathbb{E}_\mu^n\{K_{1j}\}/q_j - 1| \leqslant \delta_n/2$, thus $\mathrm{B}_{n,j}^c \subset \big\{ |\sum_{i=1}^n \zeta_{ij}|/q_j \leqslant \delta_n/2 \big\}$. Since $|\zeta_{ij}| \leqslant 2\|K\|_\infty$ and $\sum_{i=1}^n \mathbb{E}_\mu^n\{\zeta_{ij}^2\} \leqslant (1 + \delta_n) q_j \int K^2$ for n large enough, Bernstein inequality entails

$$\mathbb{P}_\mu^n\{\mathrm{B}_{n,j}^c\} \leqslant 2 \exp(-D_1 \delta_n^2 n^{2s/(2s+1)}) \tag{5.8}$$

for any $j \in \mathcal{J}_{2,n}$, where D_1 is a positive constant. Since $\varphi_s \in \Sigma(s, 1; \mathbb{R})$, we have $K \in \Sigma(\min(s,1), L_K; \mathbb{R})$ where $L_K := (\int \varphi_s)^{-1}$ if $0 < s \leqslant 1$ and $L_K := \|K'\|_\infty$ if $s > 1$. Since $\operatorname{Supp} K = [-T_s, T_s]$, we have on $\mathrm{C}_{n,j}$

$$|\bar{K}_{ij} - K_{ij}| \leqslant L_K T_s^{\min(s,1)} \Big(\frac{\delta_n}{1 - \delta_n} \Big)^{\min(s,1)} \mathbf{1}_{\mathrm{M}_{ij}} = o(1) \mathbf{1}_{\mathrm{M}_{ij}}, \tag{5.9}$$

where $\mathrm{M}_{ij} := \{|X_i - x_j| \leqslant c_s T_s (1 + \delta_n) h_j\}$. We define $\eta_{ij} := \mathbf{1}_{\mathrm{M}_{ij}} - \mathbb{P}_\mu^n\{\mathrm{M}_{ij}\}$. On $\mathrm{C}_{n,j}$ we have for n large enough $2 c_s T_s H_n^M \leqslant \delta_n$, and since $x_j \in [\tau_n, 1 - \tau_n]$,

$$x_j \leqslant 1 - \tau_n = 1 - 2 c_s T_s H_n^M \leqslant 1 - 2 c_s T_s H_j \leqslant 1 - 2 c_s T_s (1 - \delta_n) h_j \leqslant 1 - c_s T_s (1 + \delta_n) h_j$$

for n large enough. On the other hand we have similarly $x_j \geqslant c_s T_s(1+\delta_n)h_j$. Thus, since $\mu \in \Sigma_q(\nu, \varrho)$ we have

$$\Big|\frac{\mathbb{P}_\mu^n\{\mathrm{M}_{1j}\}}{(1+\delta_n)c_s h_j \mu_j} - 2T_s\Big| \leqslant \frac{1}{q}\int_{|y|\leqslant T_s} |\mu(x_j + c_s y(1+\delta_n)h_j) - \mu_j| dy = O(h_n^\nu).$$

Since $x_j \in [c_s T_s(1+\delta_n)h_j, 1-(1+\delta_n)c_s T_s h_j] \subset [c_s T_s h_j, 1 - c_s T_s h_j]$, we have for n large enough that on $\mathrm{C}_{n,j}$,

$$\Big|\frac{\mathbb{E}_{f,\mu}^n\{K_{1j}\}}{c_s H_j \mu_j} - 1\Big| \leqslant \frac{h_j}{H_j \mu_j}\int |K(y)||\mu(x_j + y c_s h_j) - \mu_j| dy + \Big|\frac{h_j}{H_j} - 1\Big| \leqslant O(h_n^\nu) + \frac{\delta_n}{1-\delta_n}.$$

Then, we obtain that on $\mathrm{C}_{n,j}$ and for n large enough,

$$\begin{aligned}\Big|\frac{\sum_{i=1}^n \bar K_{ij}}{\bar q_j} - 1\Big| &\leqslant \frac{o(1)}{\bar q_j}\Big|\sum_{i=1}^n \eta_{ij}\Big| + \frac{L_K (T_s\delta_n)^{\min(s,1)}}{(1-\delta_n)^{\min(s,1)}}\frac{\mathbb{P}_\mu^n\{\mathrm{M}_{1j}\}}{c_s H_j\mu_j} + \frac{1}{\bar q_j}\Big|\sum_{i=1}^n \zeta_{ij}\Big| + O(h_n^\nu) + \frac{\delta_n}{1-\delta_n} \\ &\leqslant \frac{o(1)}{q_j}\Big|\sum_{i=1}^n \eta_{ij}\Big| + \frac{1+o(1)}{q_j}\Big|\sum_{i=1}^n \zeta_{ij}\Big| + 2(2L_K T^{\min(s,1)+1} + 1)\delta_n^{\min(s,1)},\end{aligned}$$

and taking $L_A := 4(L_K T^{\min(s,1)+1} + 1)$, we obtain

$$\mathbb{P}_{f,\mu}^n\{\mathrm{A}_{n,j}^c \cap \mathrm{C}_{n,j}\} \leqslant \mathbb{P}_\mu^n\{\Big|\sum_{i=1}^n \eta_{ij}\Big| > \delta_n^{\min(s,1)} q_j\} + \mathbb{P}_\mu^n\{\Big|\sum_{i=1}^n \zeta_{ij}\Big| > \delta_n^{\min(s,1)} q_j/2\}.$$

Then, applying Bernstein inequality to the sum of variables η_{ij} and ζ_{ij}, $1 \leqslant i \leqslant n$, we obtain that for any $j \in \mathcal{J}_{n,2}$,

$$\mathbb{P}_\mu^n\{\mathrm{A}_{n,j}^c \cap \mathrm{C}_{n,j}\} \leqslant 2\exp(-D_2 \delta_{2,n}^2\, n^{2s/(2s+1)}), \tag{5.10}$$

where D_2 is a positive constant and $\delta_{2,n} := \delta_n^{\min(s,1)}$. We can prove

$$\mathbb{P}_\mu^n\{\mathrm{E}_{n,j}^c \cap \mathrm{C}_{n,j}\} \leqslant 2\exp(-D_3 \delta_{2,n}^2 n^{2s/(2s+1)}), \tag{5.11}$$

where D_3 is a positive constant in the same way as for the proof of (5.10), with an appropriate choice for L_E. If $I = I(x,h)$ (see (4.1)) and $\delta_{1,n} := 1 - (1+\delta_n)^{-(2s+1)}$, we define the event

$$\mathrm{N}_{n,I} := \Big\{\Big|\frac{\bar\mu_n(I)}{\mu(x)h} - 1\Big| \leqslant \delta_{1,n}\Big\}. \tag{5.12}$$

From the definitions of H_j and h_j, we obtain

$$\begin{aligned}\{(1-\delta_n)h_j < H_j\} &= \big\{(1-\delta_n)^{2s} h_j^{2s} < \log n / \big(n\bar\mu_n(I(x_j,(1-\delta_n)h_j))\big)\big\} \\ &= \Big\{\frac{\bar\mu_n(I(x_j,(1-\delta_n)h_j))}{\mu_j(1-\delta_n)h_j} \leqslant (1-\delta_n)^{-(2s+1)}\Big\},\end{aligned}$$

and then $\mathrm{N}_{n,I(x_j,(1-\delta_n)h_j)} \subset \{(1-\delta_n)h_j < H_j\}$. We can prove in the same way that on the other hand $\mathrm{N}_{n,I(x_j,(1+\delta_n)h_j)} \subset \{(1+\delta_n)h_j \geqslant H_j\}$, hence

$$\mathrm{N}_{n,I(x_j,(1-\delta_n)h_j)} \cap \mathrm{N}_{n,I(x_j,(1+\delta_n)h_j)} \subset \mathrm{C}_{n,j}. \tag{5.13}$$

If $I = I(x,h)$, we have in view of assumption D that $|\int_I \mu(t)dt - h\mu(x)| = O(h^{\nu+1})$, thus, if $Z_i := \mathbf{1}_{X_i \in I} - \int_I \mu(t)dt$, we have $\{|\sum_{i=1}^n Z_i| \leqslant n\mu(x)h\delta_{n,1}/2\} \subset \mathrm{N}_{n,I}$ for n large enough. Then, using Bernstein inequality to the sum of Z_i, $1 \leqslant i \leqslant n$, we obtain

$$\mathbb{P}_\mu^n\{\mathrm{C}_{n,j}^c\} \leqslant 2\exp(-D_C \delta_{1,n}^2 n^{2s/(2s+1)}),$$

for n large enough, where $D_C > 0$ is fixed. Using together the previous inequalities, we obtain

$$\mathbb{P}_\mu^n\{\mathcal{B}_n^c\} \leqslant \exp(-D_{\mathcal{B}} n^{s/(2s+1)}) \tag{5.14}$$

for n large enough, where $D_{\mathcal{B}} > 0$ is fixed.

Step 3: controls on $\mathcal{E}^{\Delta}_{n,f}$. We need the following lemma, which is proven below in the section.

Lemma 2. *There is an event $\mathcal{A}_n \in \mathfrak{X}_n$ such that, under assumption D,*

$$\mathbb{P}^n_{\mu}\{\mathcal{A}_n^c\} \leqslant \exp(-D_{\mathcal{A}} n^{s/(2s+1)}) \tag{5.15}$$

for n large enough, where $D_{\mathcal{A}} > 0$ and such that

$$\mathcal{A}_n \subset \mathcal{B}_n \cap \mathcal{C}_n \cap \Gamma_n, \tag{5.16}$$

where $\mathcal{B}_n$ is given by (5.7), $\mathcal{C}_n$ by lemma 1 and $\Gamma_n := \bigcap_{j \in \mathcal{J}_n} \Gamma_{n,I(x_j,h_n)}$ where $\Gamma_{n,I}$ is defined by (4.6).

In this step, we prove that for any $\varepsilon > 0$, when n is large enough, the following inequality holds:

$$\sup_{f \in \Sigma^Q(s,L)} \mathbb{P}^n_{f,\mu}\{\mathcal{E}^{\Delta}_{n,f} \mathbf{1}_{\mathcal{A}_n} > (1+\varepsilon)P\} \leqslant \exp\big(-D_{\mathcal{E}}\, \varepsilon(1 \wedge \varepsilon)(\log n)^{2s/(2s+1)}\big), \tag{5.17}$$

where $D_{\mathcal{E}} > 0$, and we prove that

$$\sup_{f \in \Sigma^Q(s,L)} \mathbb{E}^n_{f,\mu}\big\{w^2(\mathcal{E}^{\Delta}_{n,f} \mathbf{1}_{\mathcal{A}_n})\big\} = O(1). \tag{5.18}$$

We decompose the risk into

$$\mathcal{E}^{\Delta}_{n,f} = \mathcal{E}^{\Delta,1}_{n,f} + \mathcal{E}^{\Delta,2}_{n,f} + \mathcal{E}^{\Delta,3}_{n,f}, \tag{5.19}$$

where $\mathcal{E}^{\Delta,a}_{n,f} := \sup_{j \in \mathcal{J}_{a,n}} r_j^{-1} |\widehat{f}_n(x_j) - f(x_j)|$. For $a = 1$ and $a = 3$, $\mathcal{E}^{\Delta,a}_{n,f}$ is the risk at the boundaries of $[0,1]$. Since on $\mathcal{B}_n$, $\bar{Q}_j / \bar{q}_j \geqslant 1 - L_A \delta_n^{\min(s,1)} > \delta_n$ for n large enough, the denominator in (4.5) is larger than δ_n. Hence, we can decompose on $\mathcal{B}_n$ the middle risk into bias and variance terms as follows:

$$\mathcal{E}^{\Delta,2}_{n,f} \leqslant b_{n,f} + U_{n,f} + Z_n, \tag{5.20}$$

where the variance term is given by

$$Z_n := \max_{j \in \mathcal{J}_{2,n}} |Z_{n,j}|, \quad Z_{n,j} := r_j^{-1} \textstyle\sum_{i=1}^n \xi_i W_{ij},$$

with $W_{ij} := \bar{K}_{ij} / \bar{Q}_j$, and the bias terms are

$$b_{n,f} := \max_{j \in \mathcal{J}_{2,n}} |b_{n,f,j}|, \quad U_{n,f} := \max_{j \in \mathcal{J}_{2,n}} |U_{n,f,j}|,$$

where $b_{n,f,j} := \mathbb{E}^n_{f,\mu}\{B_{n,f,j} \mathbf{1}_{\mathcal{B}_n}\}$, $U_{n,f,j} := B_{n,f,j} - b_{n,f,j}$ with $B_{n,f,j} := r_j^{-1} \bar{P}_j / \bar{Q}_j$, $\bar{P}_j := \sum_{i=1}^n (f(X_i) - f(x_j)) \bar{K}_{ij}$. We use the three following inequalities: we have

$$\operatorname{limsup}_n \sup_{f \in \Sigma(s,L)} b_{n,f} \leqslant L c_s^s B(s,1), \tag{5.21}$$

where $B(s,L)$ is defined by (3.6), and there is a constant $D_U > 0$ such that for any $\varepsilon > 0$,

$$\sup_{f \in \Sigma(s,L)} \mathbb{P}^n_{f,\mu}\big\{U_{n,f} \mathbf{1}_{\mathcal{B}_n} > \varepsilon\big\} \leqslant \exp\big(-D_U\, \varepsilon(1 \wedge \varepsilon) n^{2s/(2s+1)}\big). \tag{5.22}$$

Moreover, we have for any $\varepsilon > 0$,

$$\sup_{f \in \Sigma^Q(s,L)} \mathbb{P}^n_{f,\mu}\big\{Z_n \mathbf{1}_{\mathcal{B}_n} > (1+\varepsilon) L c_s^s \|K\|_2\big\} \leqslant 2(\log n)^{2s/(2s+1)} n^{-\varepsilon/(2s+1)}. \tag{5.23}$$

We prove these inequalities below in the section. In view of (5.21), we have $b_{n,f} \leqslant (1 + 2\varepsilon) L c_s^s B(s,1)$ for n large enough, and using (3.8) we obtain

$$\{\mathcal{E}^{\Delta,2}_{n,f} \mathbf{1}_{\mathcal{B}_n} > (1+2\varepsilon)P\} \subset \{Z_n \mathbf{1}_{\mathcal{B}_n} > (1+\varepsilon) L c_s^s \|K\|_2\} \cup \{U_{n,f} \mathbf{1}_{\mathcal{B}_n} > \varepsilon L c_s^s \|K\|_2\}.$$

Then, in view of (5.22) and (5.23), it is easy to find $D_2 > 0$ such that uniformly for $f \in \Sigma^Q(s, L)$ and n large enough,

$$\mathbb{P}^n_{f,\mu}\big\{\mathcal{E}^{\Delta,2}_{n,f}\mathbf{1}_{\mathcal{B}_n} > (1+2\varepsilon)P\big\} \leqslant \exp\big(-D_2\,\varepsilon(1\wedge\varepsilon)\log n\big). \tag{5.24}$$

Now, we consider the boundary risk $\mathcal{E}^{\Delta,1}_{n,f}$ (the result is the same for $\mathcal{E}^{\Delta,3}_{n,f}$). In view of (5.4), we obtain

$$\mathcal{E}^{\Delta,1}_{n,f}\mathbf{1}_{\mathcal{C}_n} = O(\delta_n^{s/(2s+1)})(1+W^{(1)}),$$

and we have as previously $\mathbb{E}^n_{f,\mu}\{W^{(1)}\} = O((\log\log n)^{1/2})$, since $|\mathcal{J}_{1,n}| = O(\log n)$, and for any $\lambda > 0$,

$$\mathbb{P}^n_{f,\mu}\big\{W^{(1)} - \mathbb{E}^n_{f,\mu}\{W^{(1)}\} > \lambda\big\} \leqslant 2\exp(-\lambda^2/2).$$

Then, for some $D_3 > 0$, we obtain when n is large enough

$$\mathbb{P}^n_{f,\mu}\big\{\mathcal{E}^{\Delta,1}_{n,f}\mathbf{1}_{\mathcal{C}_n} > 2\varepsilon P\big\} \leqslant 2\exp\big(-D_3\varepsilon^2\delta_n^{-2s/(2s+1)}\big).$$

This inequality, together with (5.24) and the fact that $\mathcal{A}_n \subset \mathcal{B}_n \cap \mathcal{C}_n$ (see lemma 2) entails (5.17). To prove (5.18), since $w(x) \leqslant A(1+|x|^b)$, it suffices to use (5.17) and the fact that $\mathbb{E}^n_{f,\mu}\{(\mathcal{E}^{\Delta}_{n,f})^p\mathbf{1}_{\mathcal{A}_n}\} = p\int_0^{+\infty} t^{p-1}\mathbb{P}^n_{f,\mu}\{\mathcal{E}^{\Delta}_{n,f}\mathbf{1}_{\mathcal{A}_n} > t\}dt$ for any $p > 0$.

Step 4: conclusion of the proof. We need the following inequality, which is proven below in the section:

$$\sup_{f\in\Sigma^Q(s,L)} \mathbb{E}^n_{f,\mu}\big\{w^2(\mathcal{E}_{n,f})\big\} = O\big(n^{2b(1+s/(2s+1))}\big). \tag{5.25}$$

Since $w(\cdot)$ is nondecreasing, we have for any $\varepsilon > 0$

$$\begin{aligned}
\mathbb{E}^n_{f,\mu}\{w(\mathcal{E}_{n,f})\} &\leqslant \mathbb{E}^n_{f,\mu}\{w(\mathcal{E}_{n,f})\mathbf{1}_{\mathcal{A}_n\cap\mathcal{W}_n}\} + \mathbb{E}^n_{f,\mu}\{w(\mathcal{E}_{n,f})\mathbf{1}_{\mathcal{A}_n^c\cup\mathcal{W}_n^c}\}\\
&\leqslant w((1+2\varepsilon)P) + \big(\mathbb{E}^n_{f,\mu}\{w^2(\mathcal{E}_{n,f})\}\,\mathbb{P}^n_{f,\mu}\{\mathcal{A}_n^c\cup\mathcal{W}_n^c\}\big)^{1/2}\\
&+ \big(\mathbb{E}^n_{f,\mu}\big\{w^2\big((1+2\varepsilon)\mathcal{E}^{\Delta}_{n,f}\mathbf{1}_{\mathcal{A}_n}\big)\big\}\,\mathbb{P}^n_{f,\mu}\{\mathcal{E}^{\Delta}_{n,f}\mathbf{1}_{\mathcal{A}_n} > (1+\varepsilon)P\}\big)^{1/2}\\
&\leqslant w((1+2\varepsilon)P) + O\big(n^{b(1+s/(2s+1))}\exp(-(\log n)^2/4)\big)\\
&+ O\big(\exp(-D_{\mathcal{E}}\,\varepsilon(1\wedge\varepsilon)(\log n)^{2s/(2s+1)})\big) = w((1+2\varepsilon)P) + o(1),
\end{aligned}$$

where we used together lemma 2, equations (5.5), (5.17), (5.18), (5.25), and the fact that $w(\cdot)$ is continuous. Thus,

$$\operatorname{limsup}_n \sup_{f\in\Sigma^Q(s,L)} \mathbb{E}^n_{f,\mu}\{w(\mathcal{E}_{n,f})\} \leqslant w((1+2\varepsilon)P),$$

which concludes the proof of theorem 1 since ε can be chosen arbitrarily small. $\square$

The proofs of several technical inequalities have been postponed all along the proof of theorem 1. We give the proofs of these results in what follows. Since $b_{n,f}$ and $U_{n,f}$ only depend on f via its values in $[0,1]$, we have

$$\sup_{f\in\Sigma(s,L)} b_{n,f} = \sup_{f\in\Sigma(s,L;\mathbb{R})} b_{n,f}, \quad \sup_{f\in\Sigma(s,L)} U_{n,f} = \sup_{f\in\Sigma(s,L;\mathbb{R})} U_{n,f}. \tag{5.26}$$

Proof of (5.21)**.** On $\mathrm{A}_{n,j} \cap \mathrm{C}_{n,j}$ we have $(1-o(1))q_j \leqslant \bar{Q}_j \leqslant (1+o(1))q_j$ and since $\mathcal{B}_n \subset \mathrm{A}_{n,j} \cap \mathrm{C}_{n,j}$ for any $j \in \mathcal{J}_{2,n}$, we have

$$|b_{n,f,j}| = r_j^{-1} |\mathbb{E}^n_{f,\mu}\{(\bar{P}_j/\bar{Q}_j)\mathbf{1}_{\mathcal{B}_n}\}| \leqslant (1+o(1))(r_j q_j)^{-1} |\mathbb{E}^n_{f,\mu}\{\bar{P}_j \mathbf{1}_{\mathcal{B}_n}\}|.$$

Using (5.9), and introducing $\nu_{f,j}(x) := \mathbf{1}_{f(x)\geqslant f(x_j)} - \mathbf{1}_{f(x)<f(x_j)}$, $P_j := \sum_{i=1}^n (f(X_i) - f(x_j))K_{ij}$, $R_{ij} := \nu_{f,j}(X_i)(f(X_i)-f(x_j))\mathbf{1}_{\mathrm{M}_{ij}}$ and $R_j := \sum_{i=1}^n R_{ij}$, we obtain

$$\begin{aligned}
&\frac{1}{r_j q_j} |\mathbb{E}^n_{f,\mu}\{\bar{P}_j \mathbf{1}_{\mathcal{B}_n}\}| \leqslant \frac{1}{r_j q_j}\big(|\mathbb{E}^n_{f,\mu}\{P_j\}| + o(1)|\mathbb{E}^n_{f,\mu}\{R_j\}|\big) \\
&\leqslant \frac{1}{r_j \mu_j}\Big(\Big|\int (f(x_j + y c_s h_j) - f(x_j))K(y)\mu(x_j + y c_s h_j)dy\Big| \\
&+ o(1)\Big|\int_{|y|\leqslant (1+\delta_n)T_s} (f(x_j + y c_s h_j) - f(x_j))\nu_{f,j}(x_j + c_s y h_j)\mu(x_j + y c_s h_j)dy\Big|\Big),
\end{aligned}$$

and since $\mu \in \Sigma_q(\nu, \varrho)$ we have

$$\begin{aligned}
b_{n,f,j} \leqslant \frac{1+o(1)}{r_j}&\Big|\int (f(x_j + y c_s h_j) - f(x_j))K(y)dy\Big| \\
&+ \frac{o(1)}{r_j q}\int_{|y|\leqslant 2T_s} |f(x_j + y c_s h_j) - f(x_j)|dy.
\end{aligned}$$

Using (5.26) and the fact that $\Sigma(s, L; \mathbb{R})$ is invariant by translation,

$$\begin{aligned}
\sup_{f \in \Sigma(s,L;\mathbb{R})} b_{n,f,j} \leqslant (1+o(1)) \sup_{f\in\Sigma(s,L;\mathbb{R})} \max_{j \in \mathcal{J}_{2,n}} \frac{1}{r_j}\Big(\Big|\int (f(c_s h_j y) - f(0))K(y)dy\Big| \\
+ o(1)\int_{|y|\leqslant 2T} |f(c_s h_j y) - f(0)|dy\Big).
\end{aligned}$$

Now we use an argument which is known as *renormalisation*, see Donoho and Low (1992). We introduce the functional operator $\mathcal{U}_{a,b} f(\cdot) := a f(b\,\cdot)$. We have that $f \in \Sigma(s, L; \mathbb{R})$ is equivalent to $\mathcal{U}_{a,b} f \in \Sigma(s, Lab^s; \mathbb{R})$. Then, choosing $a = (L c_s^s h_j^s)^{-1}$ and $b = c_s h_j$ entails

$$\sup_{f\in\Sigma(s,L;\mathbb{R})} b_{n,f} \leqslant (1+o(1)) L c_s^s B(s,1) + o(1),$$

and the result follows. □

Proof of (5.22)**.** We recall that $U_{n,f,j} = B_{n,f,j} - \mathbb{E}^n_{f,\mu}\{B_{n,f,j}\mathbf{1}_{\mathcal{B}_n}\}$. We use the same notations as in the proof of (5.21). On $\mathcal{B}_n$ we have $(1-o(1))q_j \leqslant \bar{Q}_j \leqslant (1+o(1))q_j$, and since $\mathbb{E}^n_{f,\mu}\{\bar{P}_j^2\} = O(n^2)$, we obtain in view of lemma 2 that $|\mathbb{E}^n_{f,\mu}\{\bar{P}_j \mathbf{1}_{\mathcal{B}_n^c}\}|/(r_j q_j) = o(1)$. Then on $\mathcal{B}_n$,

$$|U_{n,f,j}| \leqslant \frac{1}{r_j q_j}\Big((1+o(1))\big|\bar{P}_j - \mathbb{E}^n_{f,\mu}\{\bar{P}_j\}\big| + o(1)\big|\mathbb{E}^n_{f,\mu}\{\bar{P}_j \mathbf{1}_{\mathcal{B}_n}\}\big|\Big) + o(1),$$

and we know from the proof of (5.21) that $|\mathbb{E}^n_{f,\mu}\{\bar{P}_j \mathbf{1}_{\mathcal{B}_n}\}|/(r_j q_j) = O(1)$, thus

$$|U_{n,f,j}| \leqslant (1+o(1))(r_j q_j)^{-1}|\bar{P}_j - \mathbb{E}^n_{f,\mu}\{\bar{P}_j\}| + o(1)$$

on $\mathcal{B}_n$. Using (5.9) we obtain that on $\mathcal{B}_n$,

$$|\bar{P}_j - \mathbb{E}^n_{f,\mu}\{\bar{P}_j\}| \leqslant |P_j - \mathbb{E}^n_{f,\mu}\{P_j\}| + o(1)|R_j - \mathbb{E}^n_{f,\mu}\{R_j\}| + o(1)|\mathbb{E}^n_{f,\mu}\{R_j\}|,$$

and again from the proof of (5.21), we know that $(r_j q_j)^{-1}|\mathbb{E}^n_{f,\mu}\{R_j\}| = O(1)$. Then, if $\lambda_j := \varepsilon r_j q_j/3$, we have for n large enough

$$\mathbb{P}^n_{f,\mu}\{|U_{n,f,j}|\mathbf{1}_{\mathcal{B}_n} > \varepsilon\} \leqslant \mathbb{P}^n_{f,\mu}\big\{|P_j - \mathbb{E}^n_{f,\mu}\{P_j\}| > \lambda_j\big\} + \mathbb{P}^n_{f,\mu}\big\{|R_j - \mathbb{E}^n_{f,\mu}\{R_j\}| > \lambda_j\big\}.$$

We use Bernstein inequality to the sum of variables $P_{ij} - \mathbb{E}^n_{f,\mu}\{P_{ij}\}$ and $R_{ij} - \mathbb{E}^n_{f,\mu}\{R_{ij}\}$, $1 \leqslant i \leqslant n$. These variables are independent and centered. We have $|P_{ij} - \mathbb{E}^n_{f,\mu}\{P_{ij}\}| \leqslant 4QK_\infty$, and with the same arguments as at the end of the proof of (5.21) we obtain

$$\textstyle\sum_{i=1}^n \mathbb{E}^n_{f,\mu}\{(P_{ij} - \mathbb{E}^n_{f,\mu}\{P_{ij}\})^2\} = O(r_j^2 q_j).$$

Then, using Bernstein inequality, we can find $D_4 > 0$ such that for n large enough

$$\mathbb{P}^n_{f,\mu}\{|P_j - \mathbb{E}^n_{f,\mu}\{P_j\}| > \lambda_j\} \leqslant 2\exp(-D_4\varepsilon(1\wedge\varepsilon)n^{s/(2s+1)}).$$

We can prove likewise the same inequality for $R_j - \mathbb{E}^n_{f,\mu}\{R_j\}$, with some different constant $D_4 > 0$. Since $|\mathcal{J}_{2,n}| \leqslant M_n$ and $M_n \exp(-Dn^{s/(2s+1)}/2) \to 0$ as $n \to +\infty$ for any $D > 0$, the result follows with an appropriate constant $D_U > 0$. □

Proof of (5.23). Conditionally on $\mathfrak{X}_n$, $Z_{n,j}$ is centered Gaussian with variance $v_j^2 := \sigma^2 r_j^{-2}\sum_{i=1}^n W_{ij}^2$. On $\mathcal{B}_n$, we have for any $j \in \mathcal{J}_{2,n}$ and n large enough

$$\sum_{i=1}^n W_{ij}^2 = \frac{S_j}{(\bar{Q}_j)^2} \leqslant (1+o(1))\frac{\|K\|_2^2}{q_j} \leqslant (1+\varepsilon)\frac{\|K\|_2^2 r_j^2}{c_s \log n},$$

where we used the definition of $h_{n,\mu}(x)$, hence $v_j^2 \leqslant (1+\varepsilon)\sigma^2\|K\|_2^2/(c_s\log n)$. Using the fact that $P(|N(0,v^2)| \geqslant \lambda) \leqslant 2\exp(-\lambda^2/(2v^2))$, we obtain

$$\mathbb{P}^n_{f,\mu}\{|Z_{n,j}|\mathbf{1}_{\mathcal{B}_n} > (1+\varepsilon)Lc_s^s\|K\|_2\} \leqslant 2\exp\Big(-\frac{(1+\varepsilon)}{2s+1}\log n\Big) = 2n^{-(1+\varepsilon)/(2s+1)},$$

and the result follows, since $|\mathcal{J}_{2,n}| \leqslant M_n \leqslant (\log n)^{2s/(2s+1)} n^{1/(2s+1)}$. □

Proof of (5.25). We prove that for any $p > 0$,

$$\sup_{f\in\Sigma^Q(s,L)} \mathbb{E}^n_{f,\mu}\{\mathcal{E}^p_{n,f}\} = O(n^{p(1+s/(2s+1))}), \tag{5.27}$$

which entails (5.25). By definition of $H_n(x)$, we have $H_n(x) \geqslant (\log n/n)^{1/(2s)}$ for any $x \in [0,1]$. Since $\|f\|_\infty \leqslant Q$, we have for any $j \in \mathcal{J}_{2,n}$,

$$|\widehat{f}_n(x_j)| \leqslant \delta_n^{-1}(n/\log n)^{1/(2s)}\|K_s\|_\infty(Q + |\bar{\xi}_n|/\sqrt{n}),$$

where $\bar{\xi}_n = \sum_{i=1}^n \xi_i/\sqrt{n}$ is standard Gaussian. Then,

$$\begin{aligned}\mathbb{E}^n_{f,\mu}\big\{|\widehat{f}_n(x_j)|^p|\mathfrak{X}_n\big\} &\leqslant \delta_n^{-p}(n/\log n)^{p/(2s)}(Q\vee 1)^p\|K_s\|_\infty^p\mathbb{E}^n_{f,\mu}\{(1+|\bar{\xi}_n|)^p|\mathfrak{X}_n\}\\ &= O(n^{p/(2s)}(\log n)^{p(1-1/(2s))}).\end{aligned}$$

We need the following lemma (its proof is given below).

Lemma 3. *For any interval $I \subset [0,1]$ and $p > 0$ we have*

$$\mathbb{E}^n_{f,\mu}\big\{|(\widehat{\theta}_I)_0|^p|\mathfrak{X}_n\big\} = O(n^{p/2}).$$

Moreover, for any $1 \leqslant m \leqslant k$, we have on $\Gamma_{n,I}$ (see section 4)

$$\mathbb{E}^n_{f,\mu}\big\{|(\widehat{\theta}_I)_m|^p|\mathfrak{X}_n\big\} = O(n^p).$$

When $j \in \mathcal{J}_{n,1}\cup\mathcal{J}_{n,3}$, we have $\widehat{f}_n(x_j) = (\widehat{\theta}_{I(x_j,t_n)})_0$, thus $\mathbb{E}^n_{f,\mu}\big\{|\widehat{f}_n(x_j)|^p|\mathfrak{X}_n\big\} = O(n^{p/2})$ in view of lemma 3. For any $j \in \mathcal{J}_n$, since $\widetilde{f}_n^{(m)}(x_j) = m!(\widehat{\theta}_{I(x_j,h_n)})_m$, we have in view of lemma 3 that on $\Gamma_{n,I(x_j,h_n)}$, $\mathbb{E}^n_{f,\mu}\big\{|\widetilde{f}_n^{(m)}(x_j)|^p|\mathfrak{X}_n\big\} = O(n^p)$ for any $1 \leqslant m \leqslant k$. Then, we have for any $\|f\|_\infty \leqslant Q$

$$\mathcal{E}_{n,f} = O((n/\log n)^{s/(2s+1)})\big(\sup_{x\in[0,1]}|\widehat{f}_n(x)| + Q\big),$$

and since

$$\sup_{x\in[0,1]}|\widehat{f}_n(x)| \leqslant \max_{j\in\mathcal{J}_n}\Big(|\widehat{f}_n(x_j)| + \big(\sum_{m=1}^{k}\frac{|\widetilde{f}_n^{(m)}(x_j)|}{m!}\big)\mathbf{1}_{\Gamma_{n,I(x_j,h_n)}}\Big) = O(n^p),$$

we obtain (5.27) and (5.25). □

Now, it remains to prove lemmas 1, 2 and 3. We need to introduce some notations. We consider the diagonal matrix $\mathbf{\Lambda}_I$ with entries $(\mathbf{\Lambda}_I)_{m,m} = \|\phi_{I,m}\|_I^{-1}$ for $0 \leqslant m \leqslant k$, where $\|\cdot\|_I^2 := \langle\cdot\,,\,\cdot\rangle_I$ (see section 4), the matrix $\mathcal{G}_I := \mathbf{\Lambda}_I\bar{\mathbf{X}}_I\mathbf{\Lambda}_I$, where $\bar{\mathbf{X}}_I$ is introduced in section 4 and the matrix $\mathcal{G}$ with entries $(\mathcal{G})_{p,q} := \chi_{p+q}/(\chi_{2p}\chi_{2q})^{1/2}$, for $0 \leqslant p,q \leqslant k$, where $\chi_m := (1+(-1)^m)/(2(m+1))$. Note that $\lambda(\mathcal{G}) > 0$, where we recall that $\lambda(M)$ is the smallest eigenvalue of a matrix M. We define the events

$$\Omega_n := \textstyle\bigcap_{j\in\mathcal{J}_n}\Omega_{n,I(x_j,h_n)} \cap \Omega_{n,I(x_j,t_n)}, \quad \mathcal{L}_n := \bigcap_{j\in\mathcal{J}_n}\mathcal{L}_{n,I(x_j,h_n)} \cap \mathcal{L}_{n,I(x_j,t_n)},$$

where $\Omega_{n,I}$, $I(x,h)$, h_n and t_n are defined in section 4 and where, if $I = I(x,h)$,

$$\mathcal{L}_{n,I} := \{|\lambda(\mathcal{G}_I) - \lambda(\mathcal{G})| \leqslant \delta_n\}.$$

For $0 \leqslant m \leqslant 2k$, an interval $I \subset [0,1]$, and $\delta > 0$, we define

$$\bar{\mathrm{D}}_{n,m,I,\delta} := \Big\{\Big|\frac{1}{\bar{\mu}_n(I)|I|^m}\int_I \phi_{I,m}\,d\bar{\mu}_n - \chi_m\Big| \leqslant \delta\Big\},$$

and $\mathcal{D}_n := \bigcap_{m=0}^{2k}\bigcap_{j\in\mathcal{J}_n}\bar{\mathrm{D}}_{n,m,I(x_j,h_n),\delta_n} \cap \bar{\mathrm{D}}_{n,m,I(x_j,t_n),\delta_n}$. For $\mathrm{N}_{n,I}$ given by (5.12), we define

$$\mathrm{N}_n := \textstyle\bigcap_{j\in\mathcal{J}_n}\mathrm{N}_{n,I(x_j,h_n)} \cap \mathrm{N}_{n,I(x_j,t_n)},$$

and we introduce

$$\mathcal{C}_n := \Omega_n \cap \mathcal{L}_n \cap \mathcal{D}_n \cap \mathrm{N}_n. \tag{5.28}$$

This event is used within lemma 1 above, where a control on its probability is given. We recall that Γ_n is defined in lemma 2.

To a vector $\theta \in \mathbb{R}^{k+1}$ we associate the polynomial $P_\theta(y) := \theta_0+\theta_1 y+\cdots+\theta_k y^k$. If $\widehat{\theta}_I$ is the solution of the system $\bar{\mathbf{X}}_I\theta = \mathbf{Y}_I$ (see section 4) for $I = I(x,h)$, we define $\widehat{f}_I(y) := P_{\widehat{\theta}_I}(y-x)$. If $V_{I,k} := \mathrm{Span}\{\phi_{I,m}; 0 \leqslant m \leqslant k\}$, we note that on $\Omega_{n,I}$, $\widehat{f}_I$ satisfies

$$\langle\widehat{f}_I\,,\,\phi\rangle_I = \langle Y\,,\,\phi\rangle_I, \quad \forall\phi \in V_{I,k}. \tag{5.29}$$

By definition, we have $\widetilde{f}_n^{(m)}(x_j) = \widehat{f}_{I(x_j,h_n)}^{(m)}(x_j)$, where $\widehat{f}_I^{(m)}$ is the derivative of order m of $\widehat{f}_I$, and $\bar{f}_n(x_j) = \widehat{f}_{I(x_j,t_n)}(x_j)$, see section 4. We recall that M_n is the cardinal of $\mathcal{J}_n$.

Proof of lemma 1. We take $I = I(x,h)$ for some $x \in [0,1]$, $h > 0$ and define the vector θ_I with coordinates $(\theta_I)_m = f^{(m)}(x)/m!$ for $0 \leqslant m \leqslant k$. Since $\bar{\mathbf{X}}_I = \mathbf{X}_I$ on $\Omega_{n,I}$, we have $\mathbf{\Lambda}_I^{-1}(\widehat{\theta}_I - \theta_I) = \mathcal{G}_I^{-1}\mathbf{\Lambda}_I\mathbf{X}_I(\widehat{\theta}_I - \theta_I)$. If $f_I(y) = P_{\theta_I}(y-x)$, we have in view of (5.29) for any $0 \leqslant m \leqslant k$:

$$(\mathbf{X}_I(\widehat{\theta}_I - \theta_I))_m = \langle\widehat{f}_I - f_I\,,\,\phi_{I,m}\rangle_I = \langle Y - f_I\,,\,\phi_{I,m}\rangle_I = \langle f - f_I\,,\,\phi_{I,m}\rangle_I + \langle\xi\,,\,\phi_{I,m}\rangle_I,$$

thus $\mathbf{X}_I(\widehat{\theta}_I - \theta_I) := \mathbf{B}_I + \mathbf{V}_I$. Since $f \in \Sigma(s,L)$,

$$(\mathbf{\Lambda}_I\mathbf{B}_I)_m \leqslant \|\phi_{I,m}\|_I^{-1}|\langle f - f_I\,,\,\phi_{I,m}\rangle_I| \leqslant \|f - f_I\|_I \leqslant Lh^s/k!,$$

then we can write

$$\mathbf{\Lambda}_I^{-1}(\widehat{\theta}_I - \theta_I) = \mathcal{G}_I^{-1}\frac{Lh^s}{k!}u + \frac{\sigma}{\sqrt{n\bar{\mu}_n(I)}}\,\mathcal{G}_I^{-1/2}\gamma_I,$$

where $u \in \mathbb{R}^{k+1}$ is such that $\|u\|_\infty \leqslant 1$ and $\gamma_I = (\sigma\sqrt{n\bar{\mu}_n(I)})^{-1}\mathcal{G}_I^{-1/2}\mathbf{\Lambda}_I\mathbf{D}_I\xi =: \mathbf{T}_I\xi$, where $\mathbf{D}_I$ is the matrix of size $n\bar{\mu}_n(I) \times (k+1)$ with entries $(\mathbf{D}_I)_{i,m} = (X_i - x)^m$, so that

$\mathbf{X}_I = (n\bar{\mu}_n(I))^{-1}\mathbf{D}_I'\mathbf{D}_I$. Since $\mathbf{T}_I'\mathbf{T}_I = \sigma^{-1}\mathbf{I}_{k+1}$, we obtain that γ_I is, conditionally on $\mathfrak{X}_n$, centered Gaussian with covariance equal to $\mathbf{I}_{k+1}$. Consider $I = I(x_j, h)$ for some $j \in \mathcal{J}_n$, $h > 0$. From the inequality $\|\cdot\|_\infty \leqslant \|\cdot\| \leqslant \sqrt{k+1}\|\cdot\|_\infty$ and since $\|\mathcal{G}_I^{-1/2}\| \leqslant \sqrt{k+1}\|\mathcal{G}_I^{-1}\|$ ($\mathcal{G}_I$ is symmetrical with entries smaller than 1 in absolute value) we get

$$
\begin{aligned}
\|\mathbf{\Lambda}_I^{-1}(\widehat{\theta}_I - \theta_I)\|_\infty &\leqslant \|\mathcal{G}_I^{-1}\frac{Lh^s}{k!}u\|_\infty + \frac{\sigma}{\sqrt{n\bar{\mu}_n(I)}}\|\mathcal{G}_I^{-1/2}\gamma_I\|_\infty \\
&\leqslant \|\mathcal{G}_I^{-1}\|(k+1)\big(Lh^s + \frac{\sigma}{\sqrt{n\bar{\mu}_n(I)}}\|\gamma_I\|_\infty\big) \\
&= \lambda^{-1}(\mathcal{G}_I)(k+1)\big(Lh^s + \frac{\sigma}{\sqrt{n\bar{\mu}_n(I)}}\max_{0\leqslant m\leqslant k}|W_{(k+1)j+m}|\big),
\end{aligned}
$$

where $W := (\gamma_{I(x_0,h)}, \ldots, \gamma_{I(x_{M_n},h)})'$. If $\mathbf{T} := (\mathbf{T}_{I(x_0,h)}, \ldots, \mathbf{T}_{I(x_{M_n},h)})'$ we have $W = \mathbf{T}\xi$, thus W is a centered Gaussian vector and for any $(k+1)j \leqslant m \leqslant (k+1)j+k$, $j \in \mathcal{J}_n$ we have

$$
\mathbb{E}_{f,\mu}^n\{W_m^2\} = (\mathbb{V}\mathrm{ar}\{W\})_{m,m} = (\mathbb{V}\mathrm{ar}\{\gamma_{I(x_j,h)}\})_{m-(k+1)j,m-(k+1)j} = 1,
$$

since $\mathbb{V}\mathrm{ar}\{\gamma_{I(x_j,h)}\} = \mathbf{I}_{k+1}$. Then, we have proved that on $\cap_{j\in\mathcal{J}_n}\Omega_{n,I(x_j,h)}$,

$$
\max_{j\in\mathcal{J}_n}\|\mathbf{\Lambda}_{I(x_j,h)}^{-1}(\widehat{\theta}_{I(x_j,h)} - \theta_{I(x_j,h)})\|_\infty \leqslant \lambda^{-1}(\mathcal{G}_{I(x_j,h)})(k+1)\big(Lh^s + \frac{\sigma}{\sqrt{n\bar{\mu}_n(I(x_j,h))}}W^M\big),
$$

where $W^M = \max_{0\leqslant m\leqslant (k+1)|\mathcal{J}_n|}|W_m|$. Since $\mathcal{C}_n \subset \mathrm{N}_n \cap \Omega_n \cap \mathcal{L}_n$, we have on $\mathcal{C}_n$ for $h = h_n$ or $h = t_n$,

$$
\max_{j\in\mathcal{J}_n}\|\mathbf{\Lambda}_{I(x_j,h)}^{-1}(\widehat{\theta}_{I(x_j,h)} - \theta_{I(x_j,h)})\|_\infty \leqslant (1+o(1))\lambda^{-1}(\mathcal{G})(k+1)\big(Lh^s + \frac{\sigma}{\sqrt{nh\mu_j}}W^M\big).
$$

Since $\mathcal{C}_n \subset \mathcal{D}_n$, we have for any $j \in \mathcal{J}_n$, $0 \leqslant m \leqslant k$,

$$
\mathcal{C}_n \subset \bar{\mathrm{D}}_{n,2m,I(x_j,h_n),\delta_n} \cap \bar{\mathrm{D}}_{n,2m,I(x_j,t_n),\delta_n},
$$

thus on $\mathcal{C}_n$, when $h = h_n$ or $h = t_n$, we clearly have

$$
(\mathbf{\Lambda}_{I(x_j,h)})_{m,m} = \|\phi_{I(x_j,h),m}\|_{I(x_j,h)}^{-1} \leqslant (1+o(1))h^{-m}\sqrt{2m+1}.
$$

Since $\widetilde{f}_n^{(m)}(x_j) - f^{(m)}(x_j) = m!\big((\widehat{\theta}_{I(x_j,h_n)})_m - (\theta_{I(x_j,h_n)})_m\big)$, it follows that on $\mathcal{C}_n$:

$$
\begin{aligned}
|\widetilde{f}_n^{(m)}(x_j) - f^{(m)}(x_j)| &\leqslant (1+o(1))\lambda^{-1}(\mathcal{G})m!\sqrt{2m+1}(k+1)h_n^{-m}(Lh_n^s + \frac{\sigma}{\sqrt{nh_n\mu_j}}W^M) \\
&= O(h_n^{s-m})(1 + (\log n)^{-1/2}W^M),
\end{aligned}
$$

thus (5.3). Inequality (5.4) is obtained similarly, and (5.2) follows from lemma 2. □

Proof of lemma 2. If $I = I(x,h)$, we define the event

$$
\mathrm{D}_{n,m,I,\delta} := \Big\{\Big|\frac{1}{\mu(x)h^{m+1}}\int_I \phi_{I,m}\,d\bar{\mu}_n - \chi_m\Big| \leqslant \delta\Big\}.
$$

Note that (5.13) entails

$$
\mathrm{D}_{n,0,I(x_j,(1-\delta_n)h_j),\delta_{1,n}} \cap \mathrm{D}_{n,0,I(x_j,(1+\delta_n)h_j),\delta_{1,n}} \subset \mathrm{C}_{n,j}, \tag{5.30}
$$

where we recall that $\delta_{1,n} := 1-(1+\delta_n)^{-(2s+1)}$. First, we prove (5.16). If $\delta_{3,n} = \delta_n/(2-\delta_n)$, we have for any interval I,

$$
\mathrm{D}_{n,m,I,\delta_{3,n}} \cap \mathrm{D}_{n,0,I,\delta_{3,n}} \subset \bar{\mathrm{D}}_{n,m,I,\delta_n}.
$$

Using the fact that $\lambda(M) = \inf_{\|x\|=1} \langle x\,,\, Mx \rangle$ for any symmetrical matrix M and since $\mathcal{G}_I$, $\mathcal{G}$, $\mathbf{X}_I$ are symmetrical, we obtain

$$\bigcap_{0 \leqslant p,q \leqslant k} \Big\{ |(\mathcal{G}_I - \mathcal{G})_{p,q}| \leqslant \frac{\delta_n}{(k+1)^2} \Big\} \subset \mathcal{L}_{n,I}, \tag{5.31}$$

and

$$\bigcap_{m=0}^{2k} \bar{\mathrm{D}}_{n,m,I,\frac{\delta_n}{(k+1)^2}} \subset \bigcap_{0 \leqslant p,q \leqslant k} \Big\{ \big|(\mathbf{X}_I - \mathbf{X})_{p,q}\big| \leqslant \frac{\delta_n}{(k+1)^2} \Big\} \subset \{ |\lambda(\mathbf{X}_I) - \lambda(\mathbf{X})| \leqslant \delta_n \}.$$

Recalling that if $I = I(x_j, h)$,

$$(\mathcal{G}_I)_{p,q} = \frac{\langle \phi_{I,p}\,,\, \phi_{I,q} \rangle_I}{\|\phi_{I,p}\|_I \|\phi_{I,q}\|_I} = \frac{\frac{1}{\mu_j h^{m+1}} \int_I \phi_{I,p+q}\, d\bar{\mu}_n}{\big(\frac{1}{\mu_j h^{m+1}} \int_I \phi_{I,2p}\, d\bar{\mu}_n \big)^{1/2} \big(\frac{1}{\mu_j h^{m+1}} \int_I \phi_{I,2q}\, d\bar{\mu}_n \big)^{1/2}},$$

we obtain for $\delta_{4,n} = \delta_n / \big((2-\delta_n)(2k+1)(k+1)^2 \big)$,

$$\mathrm{D}_{n,2p,I,\delta_{4,n}} \cap \mathrm{D}_{n,2q,I,\delta_{4,n}} \cap \mathrm{D}_{n,p+q,I,\delta_{4,n}} \subset \Big\{ |(\mathcal{G}_I - \mathcal{G})_{p,q}| \leqslant \frac{\delta_n}{(k+1)^2} \Big\},$$

thus

$$\bigcap_{m=0}^{2k} \mathrm{D}_{n,m,I,\delta_{4,n}} \subset \mathcal{L}_{n,I},$$

and clearly for n large enough, if $I = I(x_j, h_n)$ or $I = I(x_j, t_n)$,

$$\bigcap_{m=0}^{2k} \mathrm{D}_{n,m,I,\delta_{4,n}} \subset \{ |\lambda(\mathbf{X}_I) - \lambda(\mathbf{X})| \leqslant \delta_n \} \cap \Big\{ \Big| \frac{\bar{\mu}_n(I)}{|I| \mu_j} - 1 \Big| \leqslant \delta_n \Big\} \subset \Omega_{n,I}. \tag{5.32}$$

Moreover, if $I = I(x_j, h_n)$, we have on $\bar{\mathrm{D}}_{n,2m,I,\delta_n}$ for any $1 \leqslant m \leqslant k$ and n large enough,

$$\|\phi_{I,m}\|_I \geqslant (1 - o(1)) h_n^m \sqrt{2m+1} \geqslant 1/\sqrt{n}. \tag{5.33}$$

We define

$$\mathrm{D}_{n,m} := \bigcap_{j \in \mathcal{J}_n} \Big(\mathrm{D}_{n,m,I(x_j,h_n),\delta_{5,n}} \cap \mathrm{D}_{n,m,I(x_j,t_n),\delta_{5,n}} \\ \cap \mathrm{D}_{n,0,I(x_j,(1-\delta_n)h_j),\delta_{5,n}} \cap \mathrm{D}_{n,0,I(x_j,(1+\delta_n)h_j),\delta_{5,n}} \Big),$$

where $\delta_{5,n} = \delta_{4,n} \wedge \delta_{3,n} \wedge \delta_{1,n}$, $\mathrm{D}_n := \bigcap_{m=0}^{2k} \mathrm{D}_{n,m}$ and we choose

$$\mathcal{A}_n := \mathrm{D}_n \cap \mathcal{B}_n.$$

In view of (5.30), (5.31), (5.32), (5.33) we have $\mathrm{D}_n \subset \Omega_n \cap \mathcal{L}_n \cap \mathcal{D}_n \cap \Gamma_n$ and since $\mathrm{D}_{n,0,I,\delta} = \mathrm{N}_{n,I}$, we obtain $\mathrm{D}_n \subset \mathcal{C}_n \cap \Gamma_n$, thus (5.16). Now, we prove (5.15). Since $\mu \in \Sigma(\nu, \varrho)$, it is easy to see that for $I = I(x_j, h_n)$ or $I = I(x_j, t_n)$ and n large enough,

$$\Big| \mathbb{E}_\mu^n \Big\{ \frac{1}{\mu(x)|I|^{m+1}} \int_I \phi_{I,m}\, d\bar{\mu}_n \Big\} - \chi_m \Big| \leqslant \delta_{5,n}/2.$$

Then, using Bernstein inequality, we obtain for n large enough, if $h = h_n$, $h = t_n$, $h = (1-\delta_n)h_j$ or $h = (1+\delta_n)h_j$,

$$\mathbb{P}_\mu^n \{ \mathrm{D}^c_{n,m,I(x_j,h),\delta_{5,n}} \} \leqslant 2\exp(-D_4 \delta_{5,n}^2 nh) \leqslant 2\exp(-D_5 n^{s/(2s+1)}),$$

with D_4, D_5 positive constants, where we used the fact that $\delta_{5,n}^2 n^{s/(2s+1)} > 1$ for n large enough and $nh \geqslant D_6 n^{2s/(2s+1)}$. Hence, together with (5.14), we obtain (5.15). $\square$

Proof of lemma 3. If $\bar{\mu}_n(I) = 0$ we have $\widehat{\theta}_I = 0$ and the result is obvious, thus we assume $\bar{\mu}_n(I) > 0$. In this case, $\mathbf{\Lambda}_I$, $\bar{\mathbf{X}}_I$ and $\mathcal{G}_I$ are invertible, and by definition of $\widehat{\theta}_I$,

$$\widehat{\theta}_I = \mathbf{\Lambda}_I \mathbf{\Lambda}_I^{-1} \widehat{\theta}_I = \mathbf{\Lambda}_I \mathcal{G}_I^{-1} \mathbf{\Lambda}_I \bar{\mathbf{X}}_I \widehat{\theta}_I = \mathbf{\Lambda}_I \mathcal{G}_I^{-1} \mathbf{\Lambda}_I \mathbf{Y}_I = \mathbf{\Lambda}_I \mathcal{G}_I^{-1} (\mathbf{B}_I + \mathbf{V}_I),$$

where $(\mathbf{B}_I)_m = \|\phi_{I,m}\|_I^{-1} \langle f \,, \phi_{I,m} \rangle_I$ and $(\mathbf{V}_I)_m = \|\phi_{I,m}\|_I^{-1} \langle \xi \,, \phi_{I,m} \rangle_I$. Since $\|f\|_\infty \leqslant Q$ we have $|(\mathbf{B}_I)_m| = \|\phi_{I,m}\|_I^{-1} |\langle f \,, \phi_{I,m} \rangle_I| \leqslant \|f\|_I \leqslant Q$, thus $\|\mathbf{B}_I\|_\infty \leqslant Q$.

Conditionally on $\mathfrak{X}_n$, $\mathbf{V}_I$ is centered Gaussian and it is an easy computation to see that its covariance matrix is equal to $\sigma^2 (n \bar{\mu}_n(I))^{-1} \mathbf{\Lambda}_I \mathbf{X}_I \mathbf{\Lambda}_I$. Then $\mathbf{\Lambda}_I \mathcal{G}_I^{-1} \mathbf{V}_I$ is conditionally on $\mathfrak{X}_n$ centered Gaussian with covariance matrix $\sigma^2 (n \bar{\mu}_n(I))^{-1} \bar{\mathbf{X}}_I^{-1} \mathbf{X}_I \bar{\mathbf{X}}_I^{-1}$. If e_m is the canonical vector with coordinates $(e_m)_p = \mathbf{1}_{p=m}$, we have

$$|(\widehat{\theta}_I)_m| = |\langle \widehat{\theta}_I \,, e_m \rangle| = |\langle \mathbf{\Lambda}_I \mathcal{G}_I^{-1} \mathbf{B}_I \,, e_m \rangle| + \sigma \sqrt{k+1}\, \gamma,$$

where $\gamma = (\sigma \sqrt{k+1})^{-1} \langle \mathbf{\Lambda}_I \mathcal{G}_I^{-1} \mathbf{V}_I \,, e_m \rangle$. By definition, we have $\|\bar{\mathbf{X}}_I^{-1}\| = \lambda^{-1}(\bar{\mathbf{X}}_I) \leqslant \sqrt{n \bar{\mu}_n(I)}$, and clearly $\|\mathbf{X}_I\| \leqslant k+1$ and $\|\mathbf{\Lambda}_I^{-1}\| \leqslant 1$. Then, conditional on $\mathfrak{X}_n$, γ is centered Gaussian with variance

$$\frac{\langle e_m \,, \bar{\mathbf{X}}_I^{-1} \mathbf{X}_I \bar{\mathbf{X}}_I^{-1} e_m \rangle}{(k+1) n \bar{\mu}_n(I)} \leqslant \frac{\|\bar{\mathbf{X}}_I^{-1}\|^2 \|\mathbf{X}_I\|}{(k+1) n \bar{\mu}_n(I)} \leqslant 1.$$

Since $\|\mathcal{G}_I^{-1}\| \leqslant \|\mathbf{\Lambda}_I^{-1}\| \|\bar{\mathbf{X}}_I^{-1}\| \|\mathbf{\Lambda}_I^{-1}\| \leqslant \sqrt{n \bar{\mu}_n(I)} \leqslant \sqrt{n}$ and $(\mathbf{\Lambda}_I)_{0,0} = 1$, we have

$$\mathbb{E}_{f,\mu}^n \{ |(\widehat{\theta}_I)_0|^p | \mathfrak{X}_n \} \leqslant (k+1)^{p/2} n^{p/2} (Q \vee 1)^p \mathbb{E}_{f,\mu}^n \{ \big(1 + \sigma |\gamma|\big)^p | \mathfrak{X}_n \} = O(n^{p/2}),$$

for any $I \subset [0,1]$, and since $\|\mathbf{\Lambda}_I\| \leqslant \sqrt{n}$ on $\Gamma_{n,I}$, it follows that

$$\mathbb{E}_{f,\mu}^n \{ |(\widehat{\theta}_I)_m|^p | \mathfrak{X}_n \} \leqslant (k+1)^{p/2} n^{p} (Q \vee 1)^p \mathbb{E}_{f,\mu}^n \{ \big(1 + \sigma |\gamma|\big)^p | \mathfrak{X}_n \} = O(n^{p}),$$

for any $1 \leqslant m \leqslant k$. □

6. Proof of theorem 2

The proof of the lower bound consists in a classical reduction to the Bayesian risk over an hardest cubical subfamily of functions, see Korostelev (1993), Donoho (1994), Korostelev and Nussbaum (1999) and Bertin (2004b). The main difference with the former proofs is that the subfamily of functions depends on the design via the bandwidth $h_{n,\mu}(x)$, and that we work within a "small" interval I_n. We recall that φ_s is defined by (2.4) and that it has compact support $[-T_s, T_s]$. Let $h_n^I := \max_{x \in I_n} h_{n,\mu}(x)$ and

$$\Xi_n := 2 T_s c_s (2^{1/(s-k)} + 1) h_n^I.$$

If $I_n = [a_n, b_n]$, $M_n := [|I_n| \, \Xi_n^{-1}]$, we define the points

$$x_j := a_n + j \, \Xi_n, \quad j \in \mathcal{J}_n := \{1, \ldots, M_n\}. \tag{6.1}$$

We denote again $\mu_j = \mu(x_j)$, $h_j = h_{n,\mu}(x_j)$. Let us define the event

$$\mathrm{H}_{n,j} := \Big\{ \Big| \frac{1}{n c_s h_j \mu_j} \sum_{i=1}^n \varphi_s^2 \Big(\frac{X_i - x_j}{c_s h_j} \Big) - 1 \Big| \leqslant \varepsilon \Big\},$$

and $\mathrm{H}_n := \cap_{j \in \mathcal{J}_n} \mathrm{H}_{n,j}$. Together with the fact that $\|\varphi_s\|_2 = 1$, we obtain using Bernstein inequality that

$$\lim_{n \to +\infty} \mathbb{P}_\mu^n \{ \mathrm{H}_n \} = 1. \tag{6.2}$$

The subfamily of functions is defined as follows: we consider an hypercube $\Theta \subset [-1,1]^{M_n}$, and for $\theta \in \Theta$, we define

$$f(x; \theta) := \sum_{j \in \mathcal{J}_n} \theta_j f_j(x), \quad f_j(x) := L c_s^s h_j^s \varphi_s \Big(\frac{x - x_j}{c_s h_j} \Big).$$

Clearly, $f_j \in \Sigma(s,L)$. Let us show that $f(\cdot\,;\theta) \in \Sigma(s,L)$. We note that

$$\operatorname{Supp}\Big(\varphi_s\big(\frac{\cdot - x_j}{c_s h_j}\big)\Big) = \big[x_j - c_s T_s h_j,\, x_j + c_s T_s h_j\big] =: I_j.$$

If $x, y \in I_j$ then $f(x;\theta) = \theta_j f_j(x)$, $f(y;\theta) = \theta_j f_j(y)$ and the result is obvious. It suffices to consider the case $x \in I_j$ and $y \in I_{j+1}$. In this case, we have

$$\begin{aligned}
|f^{(k)}(x;\theta) - f^{(k)}(y;\theta)| &= |\theta_j f_j^{(k)}(x) - \theta_{j+1} f_{j+1}^{(k)}(y)| \\
&\leqslant |f_j^{(k)}(x) - f_j^{(k)}(x_j + c_s T_s h_j)| + |f_{j+1}^{(k)}(x_{j+1} - c_s T_s h_{j+1}) - f_{j+1}^{(k)}(y)| \\
&\leqslant L\big(|x - x_j - c_s T_s h_j|^{s-k} + |x_{j+1} - c_s T_s h_{j+1} - y|^{s-k}\big) \\
&\leqslant L\big((2c_s T_s h_j)^{s-k} + (2c_s T_s h_{j+1})^{s-k}\big) \leqslant 2L(2c_s T_s h_n^I)^{s-k}.
\end{aligned}$$

Moreover, since $x \in I_j$ and $y \in I_{j+1}$ we have $|x-y| \geqslant x_{j+1} - x_j - c_s T_s(h_j + h_{j+1}) \geqslant \Xi_n - 2c_s T_s h_n^I = 2^{1/(s-k)}(2c_s T_s h_n^I)$, and finally $|f^{(k)}(x;\theta) - f^{(k)}(y;\theta)| \leqslant L|x-y|^{s-k}$, thus $f(\cdot\,;\theta) \in \Sigma(s,L)$. For any $j \in \mathcal{J}_n$, we define the statistics

$$y_j := \frac{\sum_{i=1}^n Y_i \varphi_s(X_i)}{\sum_{i=1}^n \varphi_s^2(X_i)}.$$

Since the f_j have disjoint supports, we have that conditionally on $\mathfrak{X}_n$, the y_j are Gaussian independent with $\mathbb{E}_{f,\mu}^n\{y_j|\mathfrak{X}_n\} = \theta_j$. Since

$$v_j^2 := \mathbb{E}_{f,\mu}^n\{y_j^2|\mathfrak{X}_n\} = \frac{\sigma^2}{\sum_{i=1}^n f_j^2(X_i)},$$

we obtain that on H_n,

$$\frac{2s+1}{2(1+\varepsilon)\log n} \leqslant v_j^2 \leqslant \frac{2s+1}{2(1-\varepsilon)\log n}. \tag{6.3}$$

In the model (1.1) with $f(\cdot) = f(\cdot\,;\theta)$, conditionally on $\mathfrak{X}_n$, the likelihood function of $(Y_1,\ldots,Y_n)$ can be written on H_n in the form

$$\frac{\mathrm{d}\mathbb{P}_{f,\mu}^n}{\mathrm{d}\lambda^n}|_{\mathfrak{X}_n}(Y_1,\ldots,Y_n) = \prod_{i=1}^n g_\sigma(Y_i) \prod_{j\in\mathcal{J}_n} \frac{g_{v_j}(y_j - \theta_j)}{g_{v_j}(y_j)}, \tag{6.4}$$

where g_v is the density of $N(0,v^2)$, and λ^n is the Lebesgue measure over $\mathbb{R}^n$. This fact follows from the following computation:

$$\begin{aligned}
&\prod_{i=1}^n g_\sigma(Y_i) \prod_{j\in\mathcal{J}_n} \frac{g_{v_j}(y_j - \theta_j)}{g_{v_j}(y_j)} \\
&\quad= \frac{1}{\sigma^n(2\pi)^{n/2}} \prod_{i=1}^n \exp\big(-Y_i^2/(2\sigma^2)\big) \prod_{j\in\mathcal{J}_n} \exp\big((2\theta_j y_j - \theta_j)/(2v_j^2)\big) \\
&\quad= \frac{1}{\sigma^n(2\pi)^{n/2}} \prod_{i=1}^n \Big[\exp\Big(\frac{-Y_i^2 + \sum_{j\in\mathcal{J}_n}\big(2Y_j\theta_j f_j(X_i) - \theta_j^2 f_j(X_i)^2\big)}{2\sigma^2}\Big)\Big] \\
&\quad= \frac{1}{\sigma^n(2\pi)^{n/2}} \prod_{i=1}^n \exp\Big(-\frac{(Y_i - f(X_i;\theta))^2}{2\sigma^2}\Big) = \frac{\mathrm{d}\mathbb{P}_{f,\mu}^n}{\mathrm{d}\lambda^n}|_{\mathfrak{X}_n}(Y_1,\ldots,Y_n).
\end{aligned}$$

In the following, we denote $\Sigma = \Sigma(s,L)$ and $\mathcal{E}_{n,f,T}^I := \sup_{x\in I} r_{n,\mu}(x)^{-1}|T(x) - f(x)|$. Since $w(\cdot)$ is nondecreasing and $f(\cdot\,;\theta) \in \Sigma$ for any $\theta \in \Theta$, we have for any probability distribution $\mathcal{B}$ on Θ, by a minoration of the minimax risk by the Bayesian risk,

$$\inf_T \sup_{f\in\Sigma} \mathbb{E}_{f,\mu}^n\big\{w(\mathcal{E}_{n,f,T}^I)\big\} \geqslant w\big((1-\varepsilon)P\big) \inf_T \int_\Theta \mathbb{P}_\theta^n\big\{\mathcal{E}_{n,f,T}^I \geqslant (1-\varepsilon)P\big\}\mathcal{B}(d\theta),$$

where $\mathbb{P}^n_\theta := \mathbb{P}^n_{f(\cdot\,;\theta),\mu}$. Since by construction $f(x_j;\theta) = r_j\theta_j P$ and $x_j \in I_n$, we obtain

$$\begin{aligned}\inf_T \int_\Theta \mathbb{P}^n_\theta\{\mathcal{E}^I_{n,f,T} \geqslant (1-\varepsilon)P\}\mathcal{B}(d\theta) &\geqslant \inf_{\widehat{\theta}} \int_\Theta \int_{\mathrm{H}_n} \mathbb{P}^n_\theta\big\{\max_{j\in\mathcal{J}_n} |\widehat{\theta}_j - \theta_j| \geqslant 1-\varepsilon|\mathfrak{X}_n\big\}\mathrm{d}\mathbb{P}^n_\mu\mathcal{B}(d\theta),\\ &\geqslant \int_{\mathrm{H}_n} \inf_{\widehat{\theta}} \int_\Theta \mathbb{P}^n_\theta\big\{\max_{j\in\mathcal{J}_n} |\widehat{\theta}_j - \theta_j| \geqslant 1-\varepsilon|\mathfrak{X}_n\big\}\mathcal{B}(d\theta)\mathrm{d}\mathbb{P}^n_\mu,\end{aligned}$$

where $\inf_{\widehat{\theta}}$ is taken among any measurable vector (with respect to the observations (1.1)) in $\mathbb{R}^{M_n}$. Then, we note that theorem 2 follows from (6.2) if we prove that on H_n,

$$\sup_{\widehat{\theta}} \int_\Theta \mathbb{P}^n_\theta\big\{\max_{j\in\mathcal{J}_n} |\widehat{\theta}_j - \theta_j| < 1-\varepsilon|\mathfrak{X}_n\big\}\mathcal{B}(d\theta) = o(1). \tag{6.5}$$

We choose $\Theta := \Theta_\varepsilon^{M_n}$ where $\Theta_\varepsilon := \{-(1-\varepsilon), 1-\varepsilon\}$ and $\mathcal{B} := \bigotimes_{j\in\mathcal{J}_n} b_\varepsilon$ where $b_\varepsilon := (\delta_{-(1-\varepsilon)} + \delta_{1-\varepsilon})/2$. Note that using (6.4), the left hand side of (6.5) is smaller than

$$\int \frac{\prod_{i=1}^n g_\sigma(Y_i)}{\prod_{j\in\mathcal{J}_n} g_{v_j}(y_j)} \Big(\prod_{j\in\mathcal{J}_n} \sup_{\widehat{\theta}_j\in\mathbb{R}} \int_{\Theta_\varepsilon} \mathbf{1}_{|\widehat{\theta}_j-\theta_j|<1-\varepsilon} g_{v_j}(y_j-\theta_j)db_\varepsilon(\theta_j)\Big)dY_1\ldots dY_n,$$

and $\widehat{\theta}_j = (1-\varepsilon)\mathbf{1}_{y_j\geqslant 0} - (1-\varepsilon)\mathbf{1}_{y_j<0}$ are strategies attaining the maximum. Thus, it suffices to prove the lower bound among estimators $\widehat{\theta}$ with coordinates $\widehat{\theta}_j \in \Theta_\varepsilon$ and measurable with respect to y_j only. Since the y_j are independent with density $g_{v_j}(\cdot - \theta_j)$, the left hand side of (6.5) is smaller than

$$\begin{aligned}&\prod_{j\in\mathcal{J}_n} \max_{\widehat{\theta}_j\in\Theta_\varepsilon} \int_{\Theta_\varepsilon} \int_{\mathbb{R}} \mathbf{1}_{|\widehat{\theta}_j(u_j)-\theta_j|<1-\varepsilon}\, g_{v_j}(u_j-\theta_j)du_j\, db_\varepsilon(\theta_j)\\ &\qquad = \prod_{j\in\mathcal{J}_n}\Big(1 - \inf_{\widehat{\theta}_j\in\Theta_\varepsilon} \int_{\Theta_\varepsilon}\int_{\mathbb{R}} \mathbf{1}_{|\widehat{\theta}_j(u)-\theta_j|\geqslant 1-\varepsilon}\, g_{v_j}(u-\theta_j)du\, db_\varepsilon(\theta_j)\Big),\end{aligned}$$

and if $\Phi(x) := \int_{-\infty}^x g_1(t)dt$ and D_1 is a positive constant,

$$\begin{aligned}&\inf_{\widehat{\theta}_j\in\Theta_\varepsilon} \int_{\Theta_\varepsilon}\int_{\mathbb{R}} \mathbf{1}_{|\widehat{\theta}_j(u)-\theta_j|\geqslant 1-\varepsilon}\, g_{v_j}(u-\theta_j)du\, db_\varepsilon(\theta_j)\\ &\quad\geqslant \inf_{\widehat{\theta}_j\in\Theta_\varepsilon} \frac{1}{2}\int_{\mathbb{R}} \big(\mathbf{1}_{\widehat{\theta}_j\geqslant 0} + \mathbf{1}_{\widehat{\theta}_j<0}\big) g_{v_j}(u-(1-\varepsilon)) \wedge g_{v_j}(u+(1-\varepsilon))du\\ &\quad= \frac{1}{v_j}\int_{-\infty}^0 g_1\Big(\frac{y-(1-\varepsilon)}{v_j}\Big)du = \Phi\Big(-\frac{1-\varepsilon}{v_j}\Big) \geqslant \frac{D_1}{\sqrt{\log n}} n^{-(1-\varepsilon)^2(1+\varepsilon)/(2s+1)},\end{aligned}$$

where we used (6.3) and the fact that for $x > 0$, $\Phi(-x) = (1+o(1))\exp(-x^2/2)/(x\sqrt{2\pi})$ If $L_n := n^{-(1-\varepsilon)^2(1+\varepsilon)/(2s+1)}(\log n)^{-1/2}$, it follows that the left hand side of (6.5) is smaller than

$$(1 - D_1 L_n)^{M_n} \leqslant \exp\big(|I_n|\,\Xi_n^{-1}\log\big(1 - D_1 L_n\big)\big),$$

and if D_2 is a positive constant,

$$|I_n|\,\Xi_n^{-1} L_n = D_2 |I_n|\, n^{\varepsilon/(2s+1)} \times n^{\varepsilon^2(1-\varepsilon)/(2s+1)}(\log n)^{-1/2-1/(2s+1)} \to +\infty$$

as $n \to +\infty$, since $|I_n| n^{\varepsilon/(2s+1)} \to +\infty$, thus the theorem. □

References

Arestov, V. V. (1990). Optimal recovery of operators and related problems. *Proc. Steklov Inst. Math.*, **4** 1–20.

Bertin, K. (2004a). Asymptotically exact minimax estimation in sup-norm for anisotropic hölder classes. *Bernoulli*, **10** 873–888.

Bertin, K. (2004b). Minimax exact constant in sup-norm for nonparametric regression with random design. *J. Statist. Plann. Inference*, **123** 225–242.

Brown, L. D., Cai, T., Low, M. G. and Zhang, C.-H. (2002). Asymptotic equivalence theory for nonparametric regression with random design. *The Annals of Statistics*, **30** 688 – 707.

Brown, L. D. and Low, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *The Annals of Statistics*, **24** 2384–2398.

Donoho, D. L. (1994). Asymptotic minimax risk for sup-norm loss: Solution via optimal recovery. *Probability Theory and Related Fields*, **99** 145–170.

Donoho, D. L. and Low, M. G. (1992). Renormalization exponents and optimal pointwise rates of convergence. *The Annals of Statistics*, **20** 944–970.

Fuller, A. T. (1961). Relay control systems optimized for various performance criteria,. *Automatic and remote control*, **1**.

Gaïffas, S. (2005). Convergence rates for pointwise curve estimation with a degenerate design. *Mathematical Methods of Statistics*, **1** 1–27. Available at http://hal.ccsd.cnrs.fr/ccsd-00003086/en/ .

Korostelev, A. and Nussbaum, M. (1999). The asymptotic minimax constant for sup-norm loss in nonparametric density estimation. *Bernoulli*, **5** 1099–1118.

Korostelev, V. (1993). An asymptotically minimax regression estimator in the uniform norm up to exact contant. *Theory of Probability and its Applications*, **38** 737–743.

Ledoux, M. and Talagrand, M. (1991). *Probability in Banach spaces*, vol. 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin. Isoperimetry and processes.

Leonov, S. (1997). On the solution of an optimal recovery problem and its applications in nonparametric regression. *Mathematical Methods of Statistics*, **6** 476–490.

Leonov, S. (1999). Remarks on extremal problems in nonparametric curve estimation. *Statistics and Probability Letters*, **43** 169–178.

Lepski, O. V., Mammen, E. and Spokoiny, V. G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *The Annals of Statistics*, **25** 929–947.

Lepski, O. V. and Tsybakov, A. B. (2000). Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probability Theory and Related Fields*, **117** 17–48.

Micchelli, C. A. and Rivlin, T. J. (1977). A survey of optimal recovery. *Optimal estimation in approximation theory* 1 – 54.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics*, **10** 1040–1053.

Laboratoire de Probabilités et Modèles Aléatoires - UMR 7599, Université Paris 7 Denis-Diderot - Boîte courrier 7012, 2, place Jussieu, 75251 PARIS Cedex 05 (FRANCE)

E-mail address: gaiffas@math.jussieu.fr